

# Checking proportionality for Cox's regression model

by

Hui Hong Zhang

***Thesis***  
*for the degree of*  
***Master of Science***

*(Master i Modellering og dataanalyse)*



*Department of Mathematics*  
*Faculty of Mathematics and Natural Sciences*  
*University of Oslo*

*May 2015*

*Det matematisk- og naturvitenskapelige fakultet*  
*Universitetet i Oslo*



# Abstract

Cox's regression model is one of the most used methods in medical statistics, and the method also finds applications in other fields. The purpose of the model is to explore the relationship between the effect of covariates and the hazard rate of experiencing an event for each individual. By finding the regression coefficients in the model, one can obtain the relative risk for each covariate.

One of the crucial assumptions in Cox regression is that the hazard rates of any two individuals have to be proportional, that is, independent of time. The model is called a proportional hazards model when all covariates are fixed. A number of graphical methods and formal tests have been suggested for checking this assumption. An important method for checking this assumption is the tests based on the scaled Schoenfeld residuals (Grambsch and Therneau, 1994). Another method for model checking is the tests based on the martingale residuals which include tests based on the score process (Lin et al., 1993).

In this thesis we provide an updated and systematic review of the tests that have been proposed for checking the proportionality assumption for Cox's regression model and to study how they perform. Simulation studies are important when studying model checking. It give us the possibility to obtain information about the performance and adequacy of the model, and bias and efficiency of the estimated regression coefficient under a variety of scenarios for non-proportional hazards. Consequently, a thorough comparison of the performance of the tests under different circumstances will be performed by using both real and simulated data. The real data used for illustration is the German Breast Cancer Study Data, and we are coming to study the time to recurrence of breast cancer.



# Preface

This thesis accomplishes my degree of Master of Science under the program Modeling and Data Analysis with specialization in Statistics and Data Analysis from the University of Oslo. It corresponds to 30 credits and is written during the period of January 2015 to May 2015.

## Acknowledgements

I'd really like to take the opportunity to thank my great supervisor Ørnulf Borgan (professor in statistics at the Department of Mathematics, University of Oslo), for providing me this interesting topic as well as splendid guidance throughout this semester. I appreciate that you always showed enthusiasm on the survival analysis study and want to thank you for all those meaningful and useful conversations we had together. Without your help it would never go so smooth as it has been.

I'd also like to thank my family and especially my lovely wife Sophia for endless support and understanding.

Oslo, Norway  
May 2015

*Hui Hong Zhang*



# Contents

<b>Abstract</b>	<b>i</b>
<b>Preface</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Survival analysis</b>	<b>3</b>
2.1 Basic concepts and notations . . . . .	3
2.2 The German Breast Cancer Study: An overview . . . . .	4
2.3 Survival data and censoring . . . . .	5
2.4 Nonparametric analysis . . . . .	6
2.4.1 Nelson-Aalen estimator . . . . .	6
2.4.2 Kaplan-Meier estimator . . . . .	6
2.4.3 Illustration: The German Breast Cancer Study Data . . . . .	7
2.5 Counting processes . . . . .	8
2.6 Cox regression . . . . .	9
2.6.1 The model . . . . .	9
2.6.2 Estimation of the regression coefficients . . . . .	10
2.6.3 Estimation of the cumulative baseline hazard . . . . .	11
2.6.4 Test of hypotheses for the regression coefficients . . . . .	11
2.6.5 Illustration: The German Breast Cancer Study Data . . . . .	13
<b>3 Methods for model checking</b>	<b>17</b>
3.1 Model assumptions . . . . .	17
3.2 Model with time-dependent terms . . . . .	18
3.2.1 Case I: Time-dependent on a known function . . . . .	18
3.2.2 Case II: Time-dependent on intervals . . . . .	18
3.2.3 Illustration: The German Breast Cancer Study Data . . . . .	20
3.3 Tests based on scaled Schoenfeld residuals . . . . .	21
3.3.1 Case I: When $\beta$ is known . . . . .	22
3.3.2 Case II: When $\beta$ is unknown . . . . .	24
3.3.3 Approximation of the score test when $\beta$ is unknown . . . . .	24
3.3.4 Illustration: The German Breast Cancer Study Data . . . . .	27
3.4 Tests based on martingale residuals . . . . .	28
3.4.1 The martingale residual processes . . . . .	28
3.4.2 Tests based on the score process . . . . .	29

3.4.3	Illustration: The German Breast Cancer Study Data . . . . .	31
3.4.4	The $\chi^2$ -test based on grouped martingale residual processes . . . . .	33
3.4.5	Illustration: The German Breast Cancer Study Data . . . . .	35
<b>4</b>	<b>Model checking by simulations</b>	<b>37</b>
4.1	General considerations of survival times . . . . .	37
4.2	When the model is correctly specified . . . . .	38
4.2.1	A general procedure for simulating a data set of survival times . . . .	38
4.2.2	Check the proportionality of the correctly specified model . . . . .	39
4.3	When the model is incorrectly specified . . . . .	43
4.3.1	Model with time-varying coefficients . . . . .	43
4.3.2	Check the proportionality of the incorrectly specified model . . . . .	47
4.3.3	Another case of model with time-varying coefficients . . . . .	51
4.3.4	Check the proportionality of the incorrectly specified model . . . . .	53
<b>5</b>	<b>Concluding remarks</b>	<b>57</b>
<b>A</b>	<b>Figures</b>	<b>59</b>
	<b>Bibliography</b>	<b>71</b>



# Chapter 1

## Introduction

Survival analysis is widely applied in a variety of scientific studies, especially in medicine, demography, biology, sociology, and econometrics. It is a set of statistical methodologies for studying the occurrences of an event of interest over time for a number of subjects. The subjects under study may be humans, animals, components, etc. The event of interest in this context may be deaths, divorces, births, or failure of components, and we are most interested in the “survival time” or the failure time of the event. Survival data is a collection of survival times for the corresponding event from a study, and will often come with a mixture of complete and incomplete observations. Censoring of an incomplete observation is a common case in survival analysis, and the main reason for censoring is that the event of interest has not occurred at the closure of the study. The occurrence of an event for an individual is described by means of hazard rates and survival curves. The two well-known nonparametric methods applied to estimating the cumulative hazard rate and the survival function from censored survival data are, respectively, the Nelson-Aalen estimator and Kaplan-Meier estimator.

In survival analysis, the dependence on covariates is described by means of regression models. One of the most used and important statistical methods in medical research is Cox’s regression model, and the method also finds applications in other fields as well like demography, technical reliability, and insurance. According to a recent review (Van Noorden et al., 2014), Cox’s original paper (Cox, 1972) is the second most cited paper in the history of statistics. The purpose of the model is to explore the relationship between the effect of covariates and the hazard rate of experiencing an event for each individual. By finding the regression coefficients in the model, one can obtain the relative risk for each covariate. For instance, in general insurance, this method is quite often applied for analyzing the relative risk of accidents caused by drivers with various skill levels.

One of the crucial assumptions in Cox regression is that the hazard rates of two individuals are proportional, that is, independent of time. The model is called a proportional hazards model when all covariates are fixed. A number of graphical methods and formal tests have been suggested for checking this assumption. An example of a method is to extend the Cox regression model with one or more time-dependent terms of the form  $\gamma_j g(t)x_j$ , where  $g(t)$  is a known function and  $x_j$  is a covariate, and test the null hypothesis that  $\gamma_j = 0$  by using the likelihood ratio, score or Wald test. Other methods that have been used for checking this assumption are the tests based on the scaled Schoenfeld residuals (Grambsch and Therneau, 1994) and the tests based on the martingale residuals,

where the last one include tests based on the score process (Lin et al., 1993).

A main aim of the thesis is to give an updated and comprehensive review of the tests that have been proposed in the literature for testing proportionality for Cox's regression model and to study how they perform. Simulation studies are important when studying model checking. It give us the possibility to obtain information about the performance and adequacy of the model, and bias and efficiency of the estimated regression coefficient under a variety of scenarios for non-proportional hazards. Consequently, a thorough comparison of the performance of the tests under the variety of situations will be performed by using both real and simulated data. The real data used for illustration is the German Breast Cancer Study Data, and we are coming to study the time to recurrence of breast cancer.

The thesis is organized as follows: In Chapter 2 we introduce the basic concepts and notations in survival analysis that will be used throughout the thesis, including a brief overview of the German Breast Cancer Study Data. Counting processes and Cox regression will also be introduced and discussed. In Chapter 3 we will give an updated and comprehensive review of the tests that have been proposed in the literature for testing proportionality for Cox's regression model. How these tests have been used in practice will be discussed and applied by using the German Breast Cancer Study Data and the statistical software **R** (R Development Core Team, 2014). In Chapter 4 we use simulation to generate our own data set of survival times in different situations of proportional and non-proportional hazards. In order to get an insight of which tests are most reliable, we will consider both cases of correctly and incorrectly specified Cox model, and then perform a thorough comparison of the tests. In Chapter 5 we summarize and make some concluding remarks.

## Chapter 2

# Survival analysis

The material from this chapter is based on Sections 1.1, 1.4, 3.1, 3.2 and 4.1 in the book by Aalen, Borgan and Gjessing (2008). The purpose is to introduce some basic concepts and ideas in survival analysis. In Section 2.1 we introduce some basic concepts and notations in survival analysis. In Section 2.2 we give a brief overview of the German Breast Cancer Study, which will be used to illustrate some result in later sections. Incomplete observation of survival times due to right-censoring, and possibly also left-truncation will be explained in Section 2.3. In Section 2.4 we introduce the two important non-parametric estimators for the cumulative hazard rate and the survival function. Further, in Section 2.5 we introduce the basic concept on counting processes which will be used in later chapter. Finally, in Section 2.6 we introduce the Cox regression model, how the regression coefficients are estimated, and how to test whether a specific null hypothesis is true, which is common in survival analysis.

### 2.1 Basic concepts and notations

Survival analysis is a set of statistical concept, models and methods for studying the occurrences of an event of interest for a number of subjects. The subjects under study may be humans, animals, components of a machine, etc., while the event of interest may for instance be death, myocardial infarction, birth of a child, and failure of a component or a system. Survival analysis is much applied in different fields, especially in medicine, demography, biology, sociology, econometrics and insurance. A survival time is the time elapsed from an initiating event to a well defined endpoint where the event of interest occurs. Some more concrete examples of survival times are:

- Time to death of a patient after start of certain treatment.
- Time from entrance to discharge from a hospital.
- Time from pregnant to birth of a child.
- Time to failure of a component or a system.

The most basic concepts we need in survival analysis are *the survival function* and *the hazard rate*. The survival function  $S(t)$  can be written in the following form

$$S(t) = P(T > t), \tag{2.1}$$

which describes the probability that the event of interest has not happened by time  $t$ . The random variable  $T$  indicates the survival time. Note that at time  $t = 0$  the survival function  $S(t) = 1$ , and as time goes by it will decline to zero or a positive value as  $t$  increases. That is because over time more and more individuals will experience the event of interest, but for events that do not necessarily happen to all individuals, like divorce or getting cancer, the random variable  $T$  may be infinite.

The hazard rate  $\alpha(t)$  is defined as

$$\alpha(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t \mid T \geq t), \quad (2.2)$$

which is the instantaneous probability of the event per unit of time. That is,  $\alpha(t)dt$  is the probability that the event will occur between time  $t$  and time  $t + dt$  given that it has not occurred earlier (before time  $t$ ). Since the survival curve is a function that starts at 1 and declines over time, the hazard rate can be essentially any nonnegative function. Therefore, by integrating the hazard rate, we get the cumulative hazard rate, which is defined as

$$A(t) = \int_0^t \alpha(s)ds. \quad (2.3)$$

There is a relation between the survival function and the (cumulative) hazard rate which we obtain by using (2.2) and (2.3):

$$A'(t) = \alpha(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{S(t) - S(t + \Delta t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log\{S(t)\}.$$

Since  $S(0) = 1$ , one gets by integration that  $-\log\{S(t)\} = \int_0^t \alpha(s)ds$ , and therefore

$$S(t) = \exp \left\{ - \int_0^t \alpha(s)ds \right\} = \exp\{-A(t)\}. \quad (2.4)$$

## 2.2 The German Breast Cancer Study: An overview

The following section is a brief overview of the German Breast Cancer Study data based on a paper by Sauerbrei and Royston (1999). The data will be used for illustration.

In the period from July 1984 to December 1989, 720 patients with primary node positive breast cancer were recruited into a breast cancer study. Only 686 of the patients have complete data and are included in the data set used here. In the whole study period, patients were followed from the date of breast cancer diagnosis until recurrence or death of the disease or censoring. At the end of the study, 299 of 686 patients had a recurrence of the disease, whereas 171 of them died of breast cancer.

The German Breast Cancer Study Data contains the following eight variables which are divided into:

- Numeric coded variables: Age at diagnosis, tumor size, number of nodes involved, number of progesterone receptors, and number of estrogen receptors.
- Categorical coded variables: Menopausal status, hormone therapy, and tumor grade.

A summary of the variables for the 686 patients from the German Breast Cancer Study Data is given in Table 2.1.

**Table 2.1** A summary of the variables for the 686 patients from the German Breast Cancer Study Data: How they are coded and some summary results.

Variable	Codes/Values	Mean	Sd	Quartiles		
<u>Numeric:</u>				25%	50%	75%
Age at Diagnosis	Years	53.05	10.12	46	53	61
Tumor Size	mm	29.33	14.30	20	25	35
Number of Nodes involved	1-51	5.01	5.48	1	3	7
Number of Prog. Recep.	0-2380	110.00	202.33	7	33	132
Number of Estrg. Recep.	0-1144	96.25	152.08	8	36	114
<u>Categorical:</u>		<b>Codes: Number</b>				
Menopausal status	1=Yes, 2=No	Yes: 290; No: 396				
Hormone Therapy	1=Yes, 2=No	Yes: 440; No: 246				
Tumor Grade	1-3 (I-III)	I: 81; II: 444; III: 161				

## 2.3 Survival data and censoring

Survival data is a collection of survival times for the corresponding event from a study. Since not all individuals will experience the event of interest within the time frame of a study, the survival data will often come as a mixture of complete and incomplete observations. The incomplete observations will consequently be censored, which is common in survival analysis. For instance, we may want to use the survival data on twenty light-bulbs which are turned on simultaneously to study whether they reach the defined lifetime. If the defined lifetime is one-thousand hours and our study ends after that, for light-bulbs which have exceeded the defined lifetime, the data will be censored.

Right-censoring is the most common form of censoring. A common way of presenting right-censored data is as follows:  $n$  individuals are observed, with survival times  $T_1, T_2, \dots, T_n$ . For each  $i$ , we observe a time  $\tilde{T}_i$  which is either the true survival time  $T_i$ , or a censoring time  $C_i$ , in which case the true survival time is “to the right” of the censoring time  $C_i$ . The observation from an individual  $i$  is the pair  $(\tilde{T}_i, D_i)$  where the censoring indicator  $D_i$  is defined by

$$D_i = \begin{cases} 1, & \text{if } \tilde{T}_i = T_i \\ 0, & \text{if } \tilde{T}_i = C_i \text{ in which case it is known that } T_i > \tilde{T}_i. \end{cases} \quad (2.5)$$

In real-life studies, right-censored observations will occur when an individual withdraws from the study, is lost to follow-up or due to closure of the study. Take the German Breast Cancer Study as an example, where we are mainly interested in the recurrence-free survival time for the patients. That is how long the patients live before they either have a recurrence of the disease or die from breast cancer. Totally 387 out of 686 patients are censored. The main reason for censoring in this case would be due to closure of the study. In later sections, when we talk about a recurrence it will also consists of death caused by cancer.

A concept related to right-censoring is that of left-truncation, which may be that the individuals come under observation some time after the initiating event. For left-truncated survival data, if the time of event that truncates individuals is  $y$ , only individuals with  $T_i \geq y$  are observed. A common case of left-truncation occurs when individuals enter a

study at random ages and are followed from this *delayed entry time* until the event of interest occurs or until the event is right-censored. For instance, we may want to study the time to death for seniors at a retirement community. Those who enter the study are of random ages and corresponding to the case of left-truncation. Another example is how different types of diet (e.g. vegetarian) will get a decreased or increased chance of health disease. Participants in similar studies are usually of random ages at the entrance of the study, which is also a common case for left-truncation.

## 2.4 Nonparametric analysis

### 2.4.1 Nelson-Aalen estimator

The common nonparametric estimator applied to estimate the cumulative hazard rate from censored survival data is the Nelson-Aalen estimator. Assume that the hazard is the same for all individuals, and let  $t_1 < t_2 < \dots < t_d$  be the observed survival times for the events, that is, the  $\tilde{T}_i$ 's with the censoring indicator  $D_i = 1$  in ascending order. Then the Nelson-Aalen estimator is a sum over the observed survival times, which is given by

$$\hat{A}(t) = \sum_{t_j \leq t} \frac{1}{Y(t_j)}, \quad (2.6)$$

where  $Y(t)$  is the number at risk “just before” time  $t$ . Further, the variance of the Nelson-Aalen estimator may be estimated by

$$\hat{\sigma}^2(t) = \sum_{t_j \leq t} \frac{1}{Y(t_j)^2}. \quad (2.7)$$

It can be shown that the Nelson-Aalen estimator, evaluated at a given time  $t$ , is approximately normally distributed in large samples. Furthermore, a standard  $100(1 - \alpha)\%$  confidence interval for  $A(t)$  takes the form

$$\hat{A}(t) \pm z_{1-\alpha/2} \hat{\sigma}(t),$$

where  $z_{1-\alpha/2}$  is the  $1 - \alpha/2$  fractile of the standard normal distribution.

### 2.4.2 Kaplan-Meier estimator

The Kaplan-Meier estimator is a nonparametric method to estimate the survival function from a sample of censored survival data. It can be written in the following form

$$\hat{S}(t) = \prod_{t_j \leq t} \left\{ 1 - \frac{1}{Y(t_j)} \right\}. \quad (2.8)$$

By using Greenwood's formula, the variance of the Kaplan-Meier estimator may be estimated by

$$\hat{\tau}^2(t) = \hat{S}(t)^2 \sum_{t_j \leq t} \frac{1}{Y(t_j)\{Y(t_j) - 1\}}. \quad (2.9)$$

In large samples, evaluated at a given time  $t$ , it can be shown that the Kaplan-Meier estimator is approximately normally distributed around  $S(t)$  with the corresponding variance estimated by (2.9). Thus a standard  $100(1 - \alpha)\%$  confidence interval for  $S(t)$  is given by

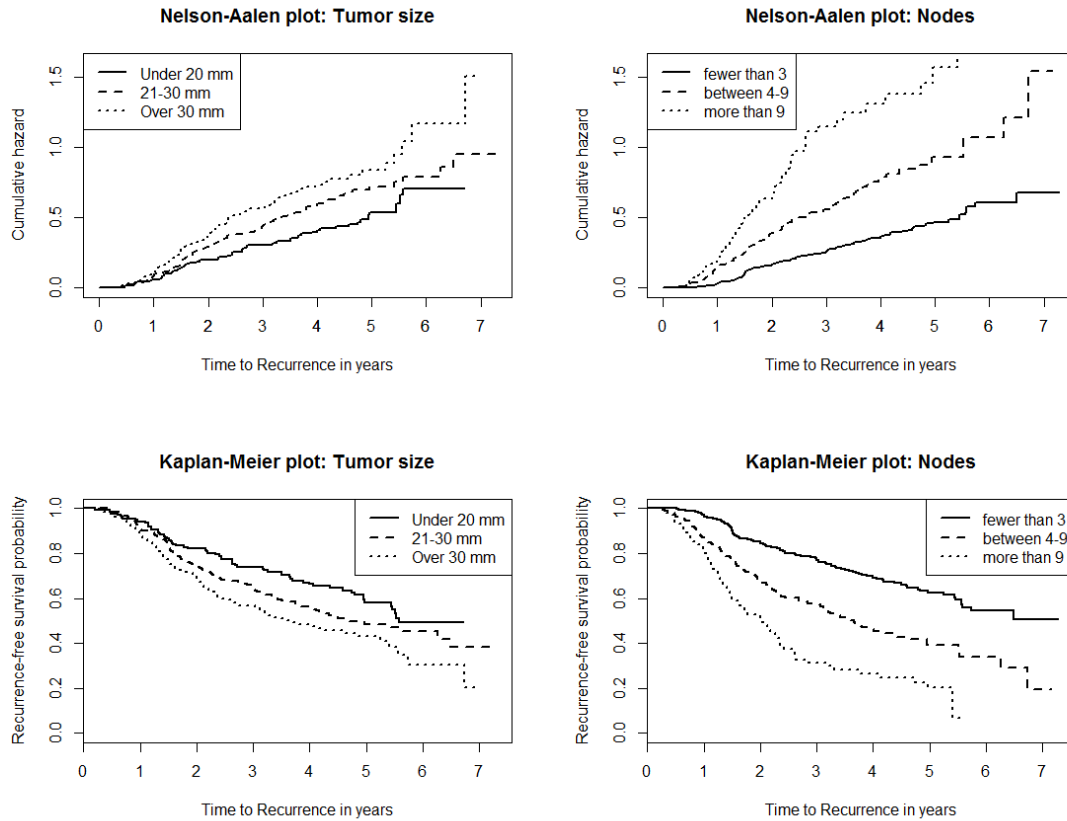
$$\hat{S}(t) \pm z_{1-\alpha/2} \hat{\tau}(t),$$

where  $z_{1-\alpha/2}$  is the  $1 - \alpha/2$  fractile of the standard normal distribution.

### 2.4.3 Illustration: The German Breast Cancer Study Data

The Nelson-Aalen and Kaplan-Meier plots in Figure 2.1 show how the size of tumor and the number of positive lymph nodes may affect the time from breast cancer diagnosis to recurrence of the disease. To ease the interpretation for the Nelson-Aalen plot, we will use, for short, “the recurrence rate” to denote “the recurrence rate of breast cancer”.

By considering the slope of the Nelson-Aalen plot for tumor size in the upper left panel, we see that the levels of the recurrence rate are not much different at the first year. After that, they are all fairly parallel. Moreover, the slopes of the plot indicates that patients with large tumor size have a bit larger recurrence rate than those with small tumor size



**Figure 2.1** Nelson-Aalen (upper panel) and Kaplan-Meier plots (lower panel) for the effect of tumor size and number of positive lymph nodes for the patients in the German Breast Cancer Study.

(smaller than 20 mm). The plots are also fairly linear which implies that the recurrence rate is approximately constant. The Kaplan-Meier plot on the left lower panel gives us much more the same conclusion as the Nelson-Aalen plot. We see that patients with tumor size smaller than 20 mm are more likely to not get a recurrence of the disease within the study compared with patients with large tumor size, but the difference is not large. More specifically, one can find that the estimated recurrence-free survival probability in three years after the breast cancer diagnosis for patients in the first and second category are 0.739 and 0.640, respectively. The corresponding estimates for patients with tumor size larger than 30 mm is only 0.570.

How the number of positive lymph nodes affects the recurrence rate can be seen from the upper right panel. According to this plot, it is clear that patients with more than 9 positive nodes have a higher recurrence rate than the other two levels. For patients with no more than 3 nodes, the recurrence rate seems to be fairly low at the first year of the study, and increased not much at the end of the study. Hence, patients that experience a recurrence from breast cancer are those with a larger number of positive lymph nodes. For the Kaplan-Meier plot in the lower right panel, the estimated recurrence-free survival probability for the third years of study are 0.771 for patients with fewer than three nodes. The corresponding estimates for the second and third levels are 0.572 and 0.314, respectively.

## 2.5 Counting processes

The focus in survival analysis is on observing the occurrence of events over time. By counting the number of events as they come along yields a counting process. For instance, one may count the number of times an individual is buying a new smartphone during the period of ten years. Or, one may also count the number of deaths from a disease in a patient group in a study.

Let  $\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_n$  be the observed event times for  $n$  individuals and denote by  $\alpha_i(t)$  the hazard rate of individual  $i$ . For a given time  $t$ , let  $N_i(t)$  be the counting process which counts the number of events that have occurred for individual  $i$  in the time interval  $[0, t]$ . The process is constant between events and jumps one unit at each event time. For survival data which contains censored event times, the counting process  $N_i(t)$  for individual  $i$  is given by

$$N_i(t) = I(\tilde{T}_i \leq t, D_i = 1); \quad i = 1, \dots, n, \quad (2.10)$$

where  $D_i$  is an censoring indicator of observing the true event time. The occurrence of future events will typically depend on “the past” for a counting process. We may then (informally) define an intensity process  $\lambda_i(t)$  for  $N_i(t)$  by

$$\lambda_i(t)dt = P(dN_i(t) = 1 \mid \text{past}) = P(t \leq \tilde{T}_i < t + dt, D_i = 1 \mid \text{past}), \quad (2.11)$$

where  $dN_i(t)$  is the number of jumps of the process in  $[t, t + dt)$  for individual  $i$ . Obviously  $\lambda_i(t) = 0$  when  $\tilde{T}_i < t$ . Consequently, the intensity process  $\lambda_i(t)$  for the counting process  $N_i(t)$  is expressed as

$$\lambda_i(t) = \alpha_i(t)Y_i(t), \quad (2.12)$$

where

$$Y_i(t) = I\{\tilde{T}_i \geq t\} \quad (2.13)$$



is an at risk indicator for individual  $i$  “just before” time  $t$ . One may obtain the aggregated counting process  $N_{\bullet}(t)$  by adding together the individual counting processes  $N_1(t)$ ,  $N_2(t), \dots, N_n(t)$ :

$$N_{\bullet}(t) = \sum_{i=1}^n N_i(t) = \sum_{i=1}^n I(\tilde{T}_i \leq t, D_i = 1). \quad (2.14)$$

The corresponding aggregated intensity process takes the form

$$\lambda_{\bullet}(t) = \sum_{i=1}^n \lambda_i(t) = \sum_{i=1}^n \alpha_i(t) Y_i(t). \quad (2.15)$$

In the case of  $\alpha_i(t) = \alpha(t)$  for all  $i$ , the intensity process is given as

$$\lambda(t) = \alpha(t) Y(t), \quad (2.16)$$

where  $Y(t) = \sum_{i=1}^n Y_i(t)$  is the number of individuals at risk “just before” time  $t$ .

## 2.6 Cox regression

In general considerations of survival analysis, a covariate in a regression model is an explanatory variable that is either of numeric or categorical type which influences the hazard rate of an individual. Usually, it is common that we deal with more than one covariate in a regression model. For instance from the German Breast Cancer Study, we are interested in the covariates Tumor Size, Hormone Therapy, Tumor Grade and Number of Positive Lymph Nodes. Further, the numeric covariates Tumor Size and Number of Positive Lymph Nodes are coded in, respectively, size in millimeter and a number of nodes counts from 1 to 51. The categorical covariate Tumor Grade and Hormone Therapy are coded in, respectively, three different grades and 1=Yes/2=No. In the following section, we will consider Cox’s regression model, which is a widely used regression model in survival analysis for censored survival data.

### 2.6.1 The model

The Cox regression model is common in survival analysis. We assume that the vector of covariates  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  for an individual  $i$  influence the hazard rate  $\alpha(t|\mathbf{x}_i)$ , which is given by the form

$$\alpha(t|\mathbf{x}_i) = \alpha_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}, \quad (2.17)$$

where  $\alpha_0(t)$  is the baseline hazard rate, the exponential function  $\exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}$  is the relative risk function, and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  is a vector of regression coefficients. Note that when all covariates are equal to zero, the relative risk function is equal to 1, such that the hazard rate corresponds to the baseline hazard.

The hazard rate ratio between two individuals, denoted 1 and 2, with the vector of covariates  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , respectively, is

$$\frac{\alpha(t|\mathbf{x}_2)}{\alpha(t|\mathbf{x}_1)} = \frac{\exp\{\boldsymbol{\beta}^T \mathbf{x}_2\}}{\exp\{\boldsymbol{\beta}^T \mathbf{x}_1\}}. \quad (2.18)$$

This ratio is constant over time when all the covariates are fixed, such that the model (2.17) in this case is the so called proportional hazards model.

If we assume that  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are exactly the same, except the  $k$ th component, which is  $x_{2k} = x_{1k} + 1$ , then (2.18) becomes

$$\frac{\alpha(t|\mathbf{x}_2)}{\alpha(t|\mathbf{x}_1)} = \exp \left\{ \boldsymbol{\beta}^T (\mathbf{x}_2 - \mathbf{x}_1) \right\} = e^{\beta_k}, \quad (2.19)$$

where  $e^{\beta_k}$  is the hazard rate ratio or the relative risk of the  $k$ th covariate. Hence, increasing the  $k$ th covariate with one unit is the same as increasing the hazard rate with a factor  $e^{\beta_k}$ , while the other covariates are kept unchanged.

### 2.6.2 Estimation of the regression coefficients

The Cox regression model (2.17) is semiparametric since the baseline hazard,  $\alpha_0(t)$ , is a nonparametric component and the relative risk function is a parametric component. To estimate the regression coefficients, ordinary likelihood methods cannot be used. Therefore, we have to look at the partial likelihood. We let  $t_1 < t_2 < \dots < t_d$  be the times when events are observed, and assume that there are no tied event times. If  $i_j$  is the index of the individual who experiences an event at  $t_j$ , then the Cox's partial likelihood for  $\boldsymbol{\beta}$  becomes

$$L(\boldsymbol{\beta}) = \prod_{t_j} \frac{\exp\{\boldsymbol{\beta}^T \mathbf{x}_{i_j}\}}{\sum_{l \in R_j} \exp\{\boldsymbol{\beta}^T \mathbf{x}_l\}}, \quad (2.20)$$

where  $R_j = \{l | Y_l(t_j) = 1\}$  is the risk set at  $t_j$  and  $Y_l(t)$  is the at risk indicator for individual  $l$  "just before" time  $t$ .

The log partial likelihood is  $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ . The maximum partial likelihood estimate  $\hat{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$  is found by maximizing (2.20) or solving the score equations

$$U_k(\boldsymbol{\beta}) = \frac{\partial}{\partial \beta_k} l(\boldsymbol{\beta}) = \sum_{t_j} \left\{ \mathbf{x}_{i_j k} - \frac{\sum_{l \in R_j} \mathbf{x}_{lk} \exp\{\boldsymbol{\beta}^T \mathbf{x}_l\}}{\sum_{l \in R_j} \exp\{\boldsymbol{\beta}^T \mathbf{x}_l\}} \right\} = 0; \quad (2.21)$$

for  $k = 1, 2, \dots, p$ . In large samples, it can be shown that  $\hat{\boldsymbol{\beta}}$  is approximately multivariate normally distributed around the true value of  $\boldsymbol{\beta}$  with a covariance matrix that may be estimated by  $\mathbf{I}(\hat{\boldsymbol{\beta}})^{-1}$  (with diagonal elements to be the estimated variances of  $\hat{\beta}$ 's, while the elements outside the diagonal are the estimated covariances between the different  $\hat{\beta}$ 's) where

$$\mathbf{I}(\boldsymbol{\beta}) = \left\{ -\frac{\partial^2}{\partial \beta_h \partial \beta_j} \log L(\boldsymbol{\beta}) \right\} \quad (2.22)$$

is the observed information matrix.

A 95% confidence interval of the relative risk (2.19) can be obtained by exponentiating the lower and upper limits of the standard 95% confidence interval for the regression coefficient,  $\hat{\beta}_k \pm 1.96 \text{se}(\hat{\beta}_k)$ , that is

$$\exp \left\{ \hat{\beta}_k \pm 1.96 \text{se}(\hat{\beta}_k) \right\}. \quad (2.23)$$

### 2.6.3 Estimation of the cumulative baseline hazard

To obtain an estimator for the cumulative baseline hazard  $A_0(t) = \int_0^t \alpha_0(u) du$ , we start out by introducing the aggregated counting process

$$N_{\bullet}(t) = \sum_{l=1}^n N_l(t).$$

By using the aggregated intensity process (2.15) with the hazard rate (2.17), its intensity process takes the form

$$\lambda_{\bullet}(t) = \left( \sum_{l=1}^n Y_l(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_l\} \right) \alpha_0(t).$$

If  $\boldsymbol{\beta}$  had been known, we could have estimated  $A_0(t)$  by the Nelson-Aalen estimator

$$\hat{A}_0(t; \boldsymbol{\beta}) = \int_0^t \frac{dN_{\bullet}(u)}{\sum_{l=1}^n Y_l(u) \exp\{\boldsymbol{\beta}^T \mathbf{x}_l\}}. \quad (2.24)$$

Since  $\boldsymbol{\beta}$  is unknown, we use the maximum partial likelihood estimator  $\hat{\boldsymbol{\beta}}$  to obtain an estimator for the cumulative baseline hazard

$$\hat{A}_0(t) = \int_0^t \frac{dN_{\bullet}(u)}{\sum_{l=1}^n Y_l(u) \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{x}_l\}} = \sum_{t_j \leq t} \frac{1}{\sum_{l \in R_j} \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{x}_l\}}. \quad (2.25)$$

The estimator (2.25) is also denoted as the Breslow estimator.

### 2.6.4 Test of hypotheses for the regression coefficients

#### Case I: When the null hypothesis is $\boldsymbol{\beta} = \boldsymbol{\beta}_0$

For Cox regression, we may want to test whether the null hypothesis  $\boldsymbol{\beta} = \boldsymbol{\beta}_0$  is true, where usually  $\boldsymbol{\beta}_0 = \mathbf{0}$ . Different test statistics could be used, but the following three types are the most widely adopted:

- The *likelihood ratio test statistic*:

$$\chi_{LR}^2 = 2\{\log L(\hat{\boldsymbol{\beta}}) - \log L(\boldsymbol{\beta}_0)\} \quad (2.26)$$

- The *score test statistic*:

$$\chi_{SC}^2 = \mathbf{U}(\boldsymbol{\beta}_0)^T \mathbf{I}(\boldsymbol{\beta}_0)^{-1} \mathbf{U}(\boldsymbol{\beta}_0) \quad (2.27)$$

where  $\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \log L(\boldsymbol{\beta})$  denotes the vector of score functions.

- The *Wald test statistic*:

$$\chi_W^2 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathbf{I}(\hat{\boldsymbol{\beta}}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \quad (2.28)$$

The test statistics (2.26), (2.27) and (2.28) are asymptotically equivalent, and under the null hypothesis, they are all approximately  $\chi^2$ -distributed with  $p$  degrees of freedom.

**Case II: When the null hypothesis is  $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$** 

Generally, one wants to test the hypothesis that  $q$  of the regression coefficients are zero (or equivalently, after a re-parameterization, that there are  $q$  linear restrictions among the coefficients). Then if  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ , the null hypothesis may be of the form of  $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$ , where usually  $\boldsymbol{\beta}_{10} = \mathbf{0}$ . Here  $\boldsymbol{\beta}_1$  is a  $q \times 1$  vector and  $\boldsymbol{\beta}_2$  is a  $(p - q) \times 1$  vector. Let  $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T)^T$  be the maximum partial likelihood estimator of  $\boldsymbol{\beta}$ . Consider the partitioned observed information matrix  $\mathbf{I} = \mathbf{I}(\boldsymbol{\beta})$  based on (2.22) expressed as

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{bmatrix}, \quad (2.29)$$

where  $\mathbf{I}_{11}$  ( $\mathbf{I}_{22}$ ) is the  $q \times q$  [ $(p - q) \times (p - q)$ ] submatrix with second partial derivatives with respect to  $\boldsymbol{\beta}_1$  ( $\boldsymbol{\beta}_2$ ), while  $\mathbf{I}_{12}$  and  $\mathbf{I}_{21}$  are submatrices defined by mixed second partial derivatives. Further, the inverse of the partitioned observed information matrix is also a partitioned matrix

$$\mathbf{I}^{-1} = \begin{bmatrix} \mathbf{I}^{11} & \mathbf{I}^{12} \\ \mathbf{I}^{21} & \mathbf{I}^{22} \end{bmatrix}. \quad (2.30)$$

It can be shown that the inverse of the submatrix  $\mathbf{I}^{11}$  is given by

$$(\mathbf{I}^{11})^{-1} = \mathbf{I}_{11} - \mathbf{I}_{12}(\mathbf{I}_{22})^{-1}\mathbf{I}_{21}. \quad (2.31)$$

Let now  $\boldsymbol{\beta}^* = [\boldsymbol{\beta}_{10}^T, \hat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}_{10})^T]^T$  be the maximum partial likelihood estimator under the null hypothesis, where  $\hat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}_{10})$  is the maximum partial likelihood estimate of  $\boldsymbol{\beta}_2$  with  $\boldsymbol{\beta}_1$  fixed at the value  $\boldsymbol{\beta}_{10}$ . Then the three test statistics takes the form:

- The *likelihood ratio test statistic*:

$$\chi_{LR}^2 = 2\{\log L(\hat{\boldsymbol{\beta}}) - \log L(\boldsymbol{\beta}^*)\} \quad (2.32)$$

- The *score test statistic*:

$$\chi_{SC}^2 = \mathbf{U}_1(\boldsymbol{\beta}^*)^T [\mathbf{I}^{11}(\boldsymbol{\beta}^*)] \mathbf{U}_1(\boldsymbol{\beta}^*) \quad (2.33)$$

where  $\mathbf{U}_1(\boldsymbol{\beta}^*)$  is the  $q \times 1$  vector of scores from  $\mathbf{U}(\boldsymbol{\beta}) = [\mathbf{U}_1(\boldsymbol{\beta})^T, \mathbf{U}_2(\boldsymbol{\beta})^T]^T$ , where  $\mathbf{U}(\boldsymbol{\beta})$  is a  $p \times 1$  vector of score functions.

- The *Wald test statistic*:

$$\chi_W^2 = (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10})^T [\mathbf{I}^{11}(\hat{\boldsymbol{\beta}})]^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) \quad (2.34)$$

where  $\mathbf{I}^{11}(\hat{\boldsymbol{\beta}})$  is the upper  $q \times q$  submatrix of (2.30) or using (2.31) directly in the formula.

We may find the score function and the corresponding observed information matrix as follows

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{t_j} \left\{ \mathbf{x}_{i_j} - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, t_j)}{\mathbf{S}^{(0)}(\boldsymbol{\beta}, t_j)} \right\}, \quad (2.35)$$

where

$$\mathbf{S}^{(0)}(\boldsymbol{\beta}, t_j) = \sum_{l \in R_j} \exp \{ \boldsymbol{\beta}^T \mathbf{x}_l \} \quad (2.36)$$

and

$$\mathbf{S}^{(1)}(\boldsymbol{\beta}, t_j) = \sum_{l \in R_j} \mathbf{x}_l \exp \{ \boldsymbol{\beta}^T \mathbf{x}_l \}. \quad (2.37)$$

The observed information matrix may be written as

$$\mathbf{I}(\boldsymbol{\beta}) = -\frac{\partial}{\partial \boldsymbol{\beta}^T} \mathbf{U}(\boldsymbol{\beta}) = \sum_{t_j} \mathbf{V}(\boldsymbol{\beta}, t_j), \quad (2.38)$$

where

$$\mathbf{V}(\boldsymbol{\beta}, t_j) = \frac{\mathbf{S}^{(2)}(\boldsymbol{\beta}, t_j)}{\mathbf{S}^{(0)}(\boldsymbol{\beta}, t_j)} - \left( \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, t_j)}{\mathbf{S}^{(0)}(\boldsymbol{\beta}, t_j)} \right)^{\otimes 2}, \quad (2.39)$$

while  $\mathbf{u}^{\otimes 2} = \mathbf{u}\mathbf{u}^T$ , and

$$\mathbf{S}^{(2)}(\boldsymbol{\beta}, t_j) = \sum_{l \in R_j} \mathbf{x}_l^{\otimes 2} \exp \{ \boldsymbol{\beta}^T \mathbf{x}_l \}. \quad (2.40)$$

The test statistics (2.32), (2.33) and (2.34) are asymptotically equivalent, and under the null hypothesis, they are all approximately  $\chi^2$ -distributed with  $q$  degrees of freedom.

### 2.6.5 Illustration: The German Breast Cancer Study Data

We start out by performing a univariate Cox regression analysis for one covariate at a time by using the German Breast Cancer Study Data. After that, we will perform a multivariate Cox regression analysis where the importance of the covariates is studied simultaneously. To ease the interpretation, we will use, “for short”, “the recurrence rate” to denote “the recurrence rate of breast cancer”. Note that the recurrence rate for the  $i$ th patient with Tumor Grade as the only covariate is given by

$$\alpha(t|\mathbf{x}_i) = \alpha_0(t) \exp \{ \beta_2 x_{i2} + \beta_3 x_{i3} \},$$

where the reference group are patients in Tumor Grade 1, while

$$x_{i2} = \begin{cases} 1, & \text{if patient } i \text{ has Tumor Grade 2} \\ 0 & \text{else} \end{cases}$$

and

$$x_{i3} = \begin{cases} 1, & \text{if patient } i \text{ has Tumor Grade 3} \\ 0 & \text{else.} \end{cases}$$

According to Table 2.1 in Section 2.2 the distribution to the numeric covariates Tumor Size, Number of Nodes, Progesterone Receptors and Estrogen Receptors are very skewed and it will be reasonable to use the base-2-logarithms transform. Since some of the observations from the covariates Progesterone Receptor and Estrogen Receptor are equal to zero, we have to add “one” to all observations before using the base-2-logarithms transform. The results of a univariate Cox regression analysis for one covariate at a time are shown in Table 2.2.

**Table 2.2** Results from a univariate Cox regression analysis for one covariate at a time. Note that  $e^{\hat{\beta}}$  is the estimated hazard ratio.

Covariates	$\hat{\beta}$	$e^{\hat{\beta}}$	$se(\hat{\beta})$	$z$ -values	$\Pr(> z )$	95%-CI for $e^{\beta}$
Age at Diagnosis	-0.004	0.996	0.006	-0.762	0.446	[0.984, 1.007]
$\log_2$ (Tumor Size)	0.384	1.468	0.089	4.323	$1.5 \times 10^{-05}$	[1.233, 1.747]
$\log_2$ (No. of Nodes)	0.376	1.457	0.044	8.574	$< 2.0 \times 10^{-16}$	[1.337, 1.588]
$\log_2$ (No. of Prog. Recep.)	-0.149	0.862	0.021	-7.231	$4.8 \times 10^{-13}$	[0.828, 0.897]
$\log_2$ (No. of Estrg. Recep.)	-0.095	0.909	0.021	-4.434	$9.2 \times 10^{-06}$	[0.998, 1.000]
Menopausal 2	0.063	1.065	0.118	0.530	0.596	[0.844, 1.342]
Hormone 2	-0.364	0.695	0.125	-2.911	0.004	[0.544, 0.888]
Tumor Grade 2	0.872	2.391	0.246	3.543	$3.9 \times 10^{-04}$	[1.476, 3.873]
Tumor Grade 3	1.154	3.170	0.262	4.411	$1.0 \times 10^{-05}$	[1.899, 5.293]

**Table 2.3** Results from a multivariate Cox regression analysis where all the estimated coefficients are significant.

Covariates	$\hat{\beta}$	$e^{\hat{\beta}}$	$se(\hat{\beta})$	$z$ -values	$\Pr(> z )$	95%-CI for $e^{\beta}$
$\log_2$ (No. of Nodes)	0.352	1.422	0.044	8.082	$6.7 \times 10^{-16}$	[1.306, 1.549]
$\log_2$ (No. of Prog. Recep.)	-0.122	0.885	0.022	-5.599	$2.2 \times 10^{-08}$	[0.849, 0.924]
Hormone 2	-0.367	0.693	0.126	-2.916	0.004	[0.541, 0.887]
Tumor Grade 2	0.524	1.688	0.252	2.081	0.037	[1.031, 2.764]
Tumor Grade 3	0.570	1.768	0.275	2.074	0.038	[1.032, 3.029]

By focusing on the relative risks  $e^{\beta}$  (or the recurrence rate ratios) in Table 2.2, we find that the estimated relative risk for Tumor Size is  $e^{0.384} \approx 1.468$ . Thus the recurrence rate for breast cancer patients are 46.8% larger per twice increase of the size of tumor. The 95% confidence interval for the relative risk does not include the value of 1, which corresponds to a significant effect of the covariate Tumor Size. Hence, the size of tumor has a positive effect on the recurrence rate for the patients. The other numerical covariates have similar interpretations.

For the categorical covariate Tumor Grade in Table 2.2, the relative risk for breast cancer patients in the second tumor grade are almost two and a half times larger than those in the reference group,  $e^{0.872} \approx 2.391$ , while the third tumor grade are more than three times larger,  $e^{1.154} \approx 3.170$ . Both covariates are significant, which implies that the risk of recurrence of breast cancer for both tumor grades are significantly different from each other. The other categorical covariates have similar interpretations.

Finally, we may fit a multivariate Cox regression model where all covariates in Table 2.2 are taken into account. At the first glance of the results from the full model fit (which is not presented here), we may observe that the  $p$ -value for the estimated coefficient for the log-transformed Number of Estrogen Receptors is 0.428, which can be omitted from the model. Then, by fitting a “new” full model without the Number of Estrogen Receptors, we may find that the estimated coefficient for the Age at Diagnosis is not significant ( $p$ -value = 0.406) and can be omitted from the next fit. Continuing in this way, we end up with the model as shown in Table 2.3 where all the estimated coefficients are significant.

From the estimated coefficients in Table 2.3, we can conclude that patients who have a large number of Progesterone Receptors and have not experience of hormone therapy, are more likely to not have a recurrence of breast cancer. Compared with patients who have a larger number of positive lymph nodes and experience of hormone therapy, they will get an increased recurrence rate. Note that in further illustrations, only the covariates in Table 2.3 will be considered.





## Chapter 3

# Methods for model checking

The following chapter is based on Section 4.1 in the book by Aalen, Borgan and Gjessing (2008), Sections 6.2 and 6.3 in the book by Therneau and Grambsch (2000), the paper by Grambsch and Therneau (1994), and the paper by Lin et al. (1993). In Section 3.1 we give a brief overview of the two crucial assumptions that have to be satisfied in Cox regression. In Section 3.2 we introduce two similar cases for models with one or more time-dependent terms, and how the proportionality may be checked by using the German Breast Cancer Study Data as an illustration. In Section 3.3 we introduce the scaled Schoenfeld residuals which is widely used for checking the proportionality assumption. Finally, in Section 3.4 we introduce the two tests based on the martingale residual processes, which include tests based on the score process (Lin et al., 1993) and the  $\chi^2$ -test based on grouped martingale residual processes (Aalen et al., 2008).

### 3.1 Model assumptions

Consider Cox's regression model with fixed covariates:

$$\alpha(t|\mathbf{x}) = \alpha_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}), \quad (3.1)$$

where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  and  $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ . There are two assumptions that have to be satisfied for Cox's regression model. The first one is the *log-linearity* for the hazard rate, which is given by taking "log" on the both sides of (3.1):

$$\log\{\alpha(t|\mathbf{x})\} = \log\{\alpha_0(t)\} + \boldsymbol{\beta}^T \mathbf{x}. \quad (3.2)$$

Hence,  $\log\{\alpha(t|\mathbf{x})\}$  has to be a linear function of the numeric covariates.

The second assumption is *proportional hazards*, which means that the hazard rates for any two individuals, denoted by 1 and 2, with the vector of covariates  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , respectively, have to be proportional. Hence, by using (3.1) we get the hazard ratio

$$\frac{\alpha(t|\mathbf{x}_2)}{\alpha(t|\mathbf{x}_1)} = \exp\{\boldsymbol{\beta}^T (\mathbf{x}_2 - \mathbf{x}_1)\}, \quad (3.3)$$

which is independent of time. This is the main assumption that will be checked by using different methods and formal tests in this thesis.

## 3.2 Model with time-dependent terms

A model is time-dependent if one or more covariates in (3.1) are dependent on a known function of time. In this case, the time-dependent model will violate the assumption of proportional hazard. In the following section, we will show how the model is defined and the assumption of proportional hazard can be checked.

### 3.2.1 Case I: Time-dependent on a known function

One way to check for proportionality is to consider a model that extends (3.1) with one or more time-dependent terms of the form  $\gamma_j g(t)x_j$ . Such a model can be written as

$$\alpha(t|\mathbf{x}) = \alpha_0(t) \exp \left\{ \boldsymbol{\beta}^T \mathbf{x} + \sum_{j=1}^q \gamma_j g(t)x_j \right\}, \quad (3.4)$$

where the  $\gamma_j$  are coefficients to be estimated, and  $g(t)$  is a known function (usually  $g(t) = \log t$ ) with  $q$  first time-dependent terms. A test for the proportional hazards assumption for the  $q$  first covariates corresponds to test the null hypothesis  $\boldsymbol{\gamma} = \mathbf{0}$ , where  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_q)^T$ , by using the likelihood ratio, score or Wald test defined in (2.32), (2.33) and (2.34), respectively.

#### Computation of the tests by using R

The likelihood ratio, score and Wald tests are implemented in the statistical software package **R** (R Development Core Team, 2014). How to use this to check the proportional hazards assumption may be explained as follows: At the beginning of the program, fit the null model (3.1) by using the `coxph` function. Then create a vector of coefficients of the form of  $(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p, 0, 0, \dots, 0)^T$ , which includes the estimated coefficients under the null model and  $q$  zeros corresponding to the null hypothesis  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_q)^T = \mathbf{0}$ . Now fit the model (3.4) by using the `coxph` function and including a number  $q$  of `tt(covariate)` arguments in the model formula. In addition, let the `init`-argument in the `coxph` function be the vector of coefficients as mentioned above. The `tt` argument is a list of time-transform functions, and by giving the argument “`tt=function(x,t,...) x*log(t)`” (exactly in this form) in the `coxph` function lets the time-transform function to be  $\log t$ . Finally, the `summary` command may be used to produce a summary of the fit, and to obtain the results from the likelihood ratio, score and Wald test. Note however that the  $p$ -values from the output will be wrong in this case, since under the null hypothesis, the tests should be  $\chi^2$ -distributed with  $q$  degrees of freedom.

### 3.2.2 Case II: Time-dependent on intervals

Another similar case related to the model (3.4) is when the model is time-dependent on intervals. How the intervals and the model are defined can be explained as follows: Let  $t_1 < t_2 < \dots < t_d$  denote the times of the observed event of interest sorted in ascending order, and then divide them into three intervals with equal number of events (e.g. by using quantiles). Then the first interval goes from 0 to  $\tau_1$ , the second interval goes from  $\tau_1$  to

$\tau_2$ , while the last interval goes from  $\tau_2$  to the end. The model can be written as

$$\alpha(t|\mathbf{x}) = \alpha_0(t) \exp \left\{ \boldsymbol{\beta}^T \mathbf{x} + \sum_{j=1}^q x_j \left[ \rho_j I(\tau_1 \leq t < \tau_2) + \kappa_j I(t \geq \tau_2) \right] \right\}, \quad (3.5)$$

where  $\rho_j$  and  $\kappa_j$  are coefficients to be estimated. The null hypothesis for the test of proportional hazard in this case corresponds to test  $\boldsymbol{\rho} = \boldsymbol{\kappa} = \mathbf{0}$ , where  $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_q)^T$  and  $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_q)^T$ , by using the likelihood ratio, score or Wald test defined in (2.32), (2.33) and (2.34), respectively.

### Computation of the tests by using R

By using the statistical software package **R** (R Development Core Team, 2014), one may first divide the observed event times into three intervals with equal number of events in each of them by using the `quantile` command. It will be needed to create two variables, where the first one catch which intervals an event or censoring occurs (with index 1, 2, or 3), while the second one is a censoring indicator for the corresponding event or censoring (1 for occurrence and 0 for censoring). Then use a `for`-loop that run over all individuals, while inside the loop, use the `if/else` statement to detect in which interval the observed event or censoring occurs. The procedure may be explained as follows:

**If an event or censoring is observed in the 1st interval:** Create a row that contains all the covariates for the corresponding individual (with start time at 0 and stop time at the event of occurrence or censoring), that is, let the index for interval be 1 and the censoring indicator be 1 or 0.

**If an event or censoring is observed in the 2nd interval:** Create two identical rows that contain all the covariates for the corresponding individual. In this case, for the first row, since the event or censoring is not observed in the first interval (between 0 and  $\tau_1$ ), let the index for the interval be 1 and the censoring indicator be 0. The second row corresponding to the observed event or censoring in the second interval (with start time at  $\tau_1$  and stop time at the event of occurrence or censoring), that is, let the index for interval be 2 and the censoring indicator be 1 or 0.

**If an event or censoring is observed in the 3rd interval:** Create three identical rows that contain all the covariates for the corresponding individual. Since the event or censoring is not observed in the first interval (between 0 and  $\tau_1$ ) or second interval (between  $\tau_1$  and  $\tau_2$ ), let the index for the interval be 1 and 2 for, respectively, the first and second row, and the censoring indicator be 0 for both. The third row corresponding to the observed event or censoring in the third interval (with start time at  $\tau_2$  and stop time at the event of occurrence or censoring), that is, let the index for interval be 3 and the censoring indicator be 1 or 0.

Now one may use the observed (start and stop) event times with the corresponding censoring indicator for all intervals to estimate the coefficients from the null model (3.1) by using the `coxph` function. Then create a vector of coefficients of the form of  $(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p, 0, 0, \dots, 0, 0, 0, \dots, 0)^T$ , which includes the estimated coefficients under the null model and a number  $2q$  of zeros corresponding to the null hypothesis  $\boldsymbol{\rho} = \boldsymbol{\kappa} = \mathbf{0}$ , where

$\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_q)^T$  and  $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_q)^T$ . Further, fit the model (3.5) by including  $q$  interaction terms between the intervals and covariates in the model formula. In addition, let the `init`-argument in the `coxph` function to be the vector of coefficients as mentioned above. Finally, the `summary` command may be used to produce a summary of the fit, and to obtain the results from the likelihood ratio, score and Wald test. Note however that the  $p$ -values from the output will be wrong in this case, since under the null hypothesis, the tests should be  $\chi^2$ -distributed with  $2q$  degrees of freedom.

### 3.2.3 Illustration: The German Breast Cancer Study Data

Consider the model (3.4) with  $q = 1$ , that is, when only one covariate is dependent on time. Then the model takes the form of

$$\alpha(t|\mathbf{x}) = \alpha_0(t) \exp \left\{ \beta_1 x_1 + \gamma_1 (\log t) x_1 + \boldsymbol{\beta}_2^T \mathbf{x}_2 \right\}, \quad (3.6)$$

where  $\beta_1$  is the coefficient corresponding to the covariate  $x_1$ ,  $\gamma_1$  is the coefficient corresponding to the time-dependent term, and  $\boldsymbol{\beta}_2^T \mathbf{x}_2$  denotes the rest of the coefficients and covariates in the model. We fit the model (3.6) based on the computation procedure in Subsection 3.2.1 with one time-dependent covariate at a time. The results for the time-dependent term from five fits are summarized in Table 3.1.

The test of the proportional hazards assumption corresponds to testing the null hypothesis of  $\gamma_1 = 0$ . According to the  $p$ -values from Table 3.1, there are significant effect of the time-dependent term for the log-transformed Progesterone Receptors and the third tumor grade. As time goes by, we can conclude that the third tumor grade have a negative effect of the hazard, which also means that the effect for a recurrence of breast cancer or death are decreasing over time. On the other hand, the estimated coefficient for the time-dependent term log-transformed Progesterone Receptors is positive, which means that the recurrence rate is increasing over time for breast cancer patients.

Now, consider the model (3.5) with  $q = 1$ , that is, when only one covariate is time-dependent on intervals. Then the model is given by

$$\alpha(t|\mathbf{x}) = \alpha_0(t) \exp \left\{ \beta_1 x_1 + \rho_1 I(\tau_1 \leq t < \tau_2) x_1 + \kappa_1 I(t \geq \tau_2) x_1 + \boldsymbol{\beta}_2^T \mathbf{x}_2 \right\}, \quad (3.7)$$

where  $\rho_1$  and  $\kappa_1$  are the coefficients corresponding to the observed event times in the second and third interval, respectively, and  $\boldsymbol{\beta}_2^T \mathbf{x}_2$  denotes the rest of the coefficients and covariates in the model. The results shown in Table 3.2 are based on the computation procedure in Subsection 3.2.2 with one time-dependent term at a time.

**Table 3.1** Results based on the model (3.6) with one time-dependent term at a time.

Time-dependent term	$\hat{\gamma}_1$	se( $\hat{\gamma}_1$ )	$\chi_{LR}^2$	$\chi_{SC}^2$	$\chi_W^2$	Pr( $>\chi_{SC,df=1}^2$ )
log $t$ : log <sub>2</sub> (No. of Nodes)	-0.102	0.069	2.219	2.199	2.193	0.138
log $t$ : log <sub>2</sub> (No. of Prog. R.)	0.066	0.033	4.166	4.101	4.072	0.043
log $t$ : Hormone 2	0.098	0.191	0.266	0.264	0.264	0.607
log $t$ : Tumor Grade 2	0.347	0.194	3.255	3.233	3.202	0.072
log $t$ : Tumor Grade 3	-0.619	0.215	8.753	8.583	8.352	0.003

**Table 3.2** Results based on the model (3.7) with one time-dependent term at a time.

Time-dependent term	$\hat{\rho}_1$	$\hat{\kappa}_1$	$\chi_{LR}^2$	$\chi_{SC}^2$	$\chi_W^2$	$\Pr(>\chi_{SC,df=2}^2)$
factor(int.) : $\log_2(\text{No. of Nodes})$	-0.125	-0.203	3.620	3.598	3.589	0.165
factor(int.) : $\log_2(\text{No. of Prog. R.})$	0.078	0.110	4.938	4.892	4.869	0.087
factor(int.) : Hormone 2	0.305	0.198	0.966	0.958	0.956	0.619
factor(int.) : Tumor Grade 2	0.560	0.404	3.720	3.739	3.715	0.154
factor(int.) : Tumor Grade 3	-0.838	-0.768	8.825	8.947	8.810	0.011

The proportional hazards assumption in this case corresponds to check the null hypothesis of  $\rho_1 = \kappa_1 = 0$ . By looking at the  $p$ -values from the score test, the time-dependent term for the third tumor grade is significant, while the other terms are non-significant which implies the independence of time. The relative risk for patients with the third tumor grade in the second and third time intervals are  $e^{-0.838} \approx 0.433$  and  $e^{-0.768} \approx 0.464$ , respectively. Hence, for patients in the second or third time interval, the effect of a recurrence of breast cancer or death are decreasing over time. Note that for those who are in the third time interval, the effect of a recurrence of breast cancer or death are approximate 7.3% ( $e^{-0.768}/e^{-0.838} \approx 1.073$ ) larger than those who are in the second time interval.

### 3.3 Tests based on scaled Schoenfeld residuals

A test of the proportional hazards assumption may be performed by using the scaled Schoenfeld residuals. In the following section, we will explain how this method is defined based on a paper by Grambsch and Therneau (1994) and the book by Therneau and Grambsch (2000). From now on, the Cox regression model with time-dependent terms added for *all* covariates becomes

$$\begin{aligned}
\alpha(t|\mathbf{x}) &= \alpha_0(t) \exp \left\{ \boldsymbol{\beta}^T \mathbf{x} + \sum_{j=1}^p \gamma_j g(t) x_j \right\} \\
&= \alpha_0(t) \exp \left\{ \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{G}(t) \mathbf{x} \right\} \\
&= \alpha_0(t) \exp \left\{ [\boldsymbol{\beta}^T + \boldsymbol{\gamma}^T \mathbf{G}(t)] \mathbf{x} \right\} \\
&= \alpha_0(t) \exp \left\{ \boldsymbol{\beta}(t)^T \mathbf{x} \right\},
\end{aligned} \tag{3.8}$$

where  $\boldsymbol{\beta}(t)^T = \boldsymbol{\beta}^T + \boldsymbol{\gamma}^T \mathbf{G}(t)$ . Note that  $\mathbf{G}(t)$  is a  $p \times p$  diagonal matrix with  $g(t)$  as the diagonal elements, and  $\mathbf{G}(t)^T = \mathbf{G}(t)$ . Let  $t_1 < t_2 < \dots < t_d$  be the event times, that is, the  $\tilde{T}_i$ 's with the censoring indicator  $D_i = 1$  in ascending order. Then the  $p \times 1$  vector of Schoenfeld residuals at time  $t_j$  are defined as

$$\mathbf{s}_j(\boldsymbol{\beta}) = \mathbf{x}_{i_j} - \mathbf{E}(\boldsymbol{\beta}, t_j); \quad \text{for } j = 1, 2, \dots, d, \tag{3.9}$$

where  $\mathbf{x}_{i_j}$  is the covariate vector of the individual experiencing the event at time  $t_j$ . The corresponding weighted mean and covariance matrix of the covariate vector at time  $t_j$  are, respectively,

$$\mathbf{E}(\boldsymbol{\beta}, t_j) = \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, t_j)}{\mathbf{S}^{(0)}(\boldsymbol{\beta}, t_j)} \tag{3.10}$$

and  $\mathbf{V}(\boldsymbol{\beta}, t_j)$  which is defined in (2.39). The definition of  $\mathbf{S}^{(0)}(\boldsymbol{\beta}, t_j)$ ,  $\mathbf{S}^{(1)}(\boldsymbol{\beta}, t_j)$  and  $\mathbf{S}^{(2)}(\boldsymbol{\beta}, t_j)$  are respectively (2.36), (2.37) and (2.40).

### 3.3.1 Case I: When $\boldsymbol{\beta}$ is known

Let us first assume that  $\boldsymbol{\beta}$  is known. Then the Schoenfeld residuals in (3.9) can be written as

$$\mathbf{s}_j(\boldsymbol{\beta}) = \left( \mathbf{x}_{i_j} - \mathbf{E}\{\boldsymbol{\beta}(t_j), t_j\} \right) + \left( \mathbf{E}\{\boldsymbol{\beta}(t_j), t_j\} - \mathbf{E}(\boldsymbol{\beta}, t_j) \right), \quad (3.11)$$

where  $\boldsymbol{\beta}(t_j) = \boldsymbol{\beta} + \mathbf{G}_j \boldsymbol{\gamma}$  with  $\mathbf{G}_j = \mathbf{G}(t_j)$ . Note in (3.11) that the first parentheses is a mean zero random variable, while the second parentheses is the difference between the weighted means under the true and null models. Then by taking the expectation of (3.11), we get

$$\mathbf{E}[\mathbf{s}_j(\boldsymbol{\beta})] = \mathbf{E}\{\boldsymbol{\beta}(t_j), t_j\} - \mathbf{E}(\boldsymbol{\beta}, t_j). \quad (3.12)$$

By using a one-term Taylor's expansion about  $\boldsymbol{\beta}(t_j) = \boldsymbol{\beta}$  to expand the first term on the right-hand side of (3.12), we get

$$\begin{aligned} \mathbf{E}\{\boldsymbol{\beta}(t_j), t_j\} &\approx \mathbf{E}(\boldsymbol{\beta}, t_j) + \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{E}(\boldsymbol{\beta}, t_j) \{\boldsymbol{\beta}(t_j) - \boldsymbol{\beta}\} \\ &= \mathbf{E}(\boldsymbol{\beta}, t_j) + \mathbf{V}(\boldsymbol{\beta}, t_j) \{\boldsymbol{\beta} + \mathbf{G}_j \boldsymbol{\gamma} - \boldsymbol{\beta}\} \\ &= \mathbf{E}(\boldsymbol{\beta}, t_j) + \mathbf{V}(\boldsymbol{\beta}, t_j) \mathbf{G}_j \boldsymbol{\gamma}, \end{aligned}$$

where  $\boldsymbol{\beta}(t_j) = \boldsymbol{\beta} + \mathbf{G}_j \boldsymbol{\gamma}$  is inserted in the second equality, and  $\frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{E}(\boldsymbol{\beta}, t_j) = \mathbf{V}(\boldsymbol{\beta}, t_j)$  follows from (2.38). Then, we obtain

$$\begin{aligned} \mathbf{E}[\mathbf{s}_j(\boldsymbol{\beta})] &\approx \mathbf{E}(\boldsymbol{\beta}, t_j) + \mathbf{V}(\boldsymbol{\beta}, t_j) \mathbf{G}_j \boldsymbol{\gamma} - \mathbf{E}(\boldsymbol{\beta}, t_j) \\ &= \mathbf{V}(\boldsymbol{\beta}, t_j) \mathbf{G}_j \boldsymbol{\gamma}. \end{aligned} \quad (3.13)$$

Further, let the scaled Schoenfeld residuals be

$$\mathbf{s}_j^*(\boldsymbol{\beta}) = \mathbf{V}^{-1}(\boldsymbol{\beta}, t_j) \mathbf{s}_j(\boldsymbol{\beta}). \quad (3.14)$$

Then, the expectation and covariance matrix of the scaled Schoenfeld residuals are

$$\mathbf{E}[\mathbf{s}_j^*(\boldsymbol{\beta})] \approx \mathbf{G}_j \boldsymbol{\gamma} \quad (3.15)$$

and

$$\begin{aligned} \text{var}[\mathbf{s}_j^*(\boldsymbol{\beta})] &= \mathbf{V}^{-1}(\boldsymbol{\beta}, t_j) \mathbf{V}[\boldsymbol{\beta}(t_j), t_j] \left[ \mathbf{V}^{-1}(\boldsymbol{\beta}, t_j) \right]^T \\ &\approx \mathbf{V}^{-1}(\boldsymbol{\beta}, t_j). \end{aligned} \quad (3.16)$$

This suggests a weighted multivariate linear model for the scaled Schoenfeld residuals. Let  $\mathbf{V}_j = \mathbf{V}(\boldsymbol{\beta}, t_j)$  for ease of the notation, and note that the  $\mathbf{s}_j(\boldsymbol{\beta})$ 's are uncorrelated. We want to estimate  $\boldsymbol{\gamma}$  by using the weighted multivariate generalized least squares method, that is, by minimizing

$$\begin{aligned} Q(\boldsymbol{\gamma}) &= \sum_{t_j} \left[ \mathbf{s}_j^*(\boldsymbol{\beta}) - \mathbf{G}_j \boldsymbol{\gamma} \right]^T \left[ \mathbf{V}_j^{-1} \right]^{-1} \left[ \mathbf{s}_j^*(\boldsymbol{\beta}) - \mathbf{G}_j \boldsymbol{\gamma} \right] \\ &= \sum_{t_j} \left[ \mathbf{s}_j^*(\boldsymbol{\beta}) - \mathbf{G}_j \boldsymbol{\gamma} \right]^T \mathbf{V}_j \left[ \mathbf{s}_j^*(\boldsymbol{\beta}) - \mathbf{G}_j \boldsymbol{\gamma} \right], \end{aligned}$$

where we used in the second equality that the inverse of an inverse  $p \times p$  matrix is the matrix itself. By differentiating  $\mathbf{Q}(\boldsymbol{\gamma})$  with respect to  $\boldsymbol{\gamma}$ , set the equation equal to 0, and then solve with respect to  $\boldsymbol{\gamma}$ , we get

$$\begin{aligned}\hat{\boldsymbol{\gamma}} &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1} \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{s}_j^*(\boldsymbol{\beta}) \\ &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1} \sum_{t_j} \mathbf{G}_j \mathbf{s}_j(\boldsymbol{\beta}),\end{aligned}\tag{3.17}$$

which is an estimate of  $\boldsymbol{\gamma}$ . Further, the covariance matrix of  $\hat{\boldsymbol{\gamma}}$  is expressed as

$$\begin{aligned}\text{var}(\hat{\boldsymbol{\gamma}}) &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1} \left( \sum_{t_j} \mathbf{G}_j \left[ \text{var}[\mathbf{s}_j(\boldsymbol{\beta})] \right] \mathbf{G}_j \right) \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1} \\ &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1} \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right) \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1} \\ &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1}.\end{aligned}\tag{3.18}$$

It can be shown by plugging (2.39) into (2.38) with

$$\mathbf{S}^{(i)}(\boldsymbol{\beta}, \boldsymbol{\gamma}, t_j) = \sum_{l \in R_j} \frac{\partial^i}{(\partial \boldsymbol{\gamma}^T)^i} \exp\{\boldsymbol{\beta}^T \mathbf{x}_l + \boldsymbol{\gamma}^T \mathbf{G}_j \mathbf{x}_l\}; \quad \text{for } i = 0, 1, 2,\tag{3.19}$$

and solve under the null hypothesis of  $\boldsymbol{\gamma} = \mathbf{0}$ , where  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)^T$ , that the observed information matrix is given by

$$\mathbf{I} = \mathbf{I}(\boldsymbol{\beta}) = \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j.\tag{3.20}$$

Then by taking the inverse of the observed information matrix, we just get

$$\mathbf{I}^{-1} = \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1},\tag{3.21}$$

which is the covariance matrix of  $\hat{\boldsymbol{\gamma}}$ . Note that  $(\mathbf{I}^{-1})^T = \mathbf{I}^{-1}$ . A test-statistic of the null hypothesis  $\boldsymbol{\gamma} = \mathbf{0}$  is given by

$$\begin{aligned}\chi_{\hat{\boldsymbol{\gamma}}}^2 &= \hat{\boldsymbol{\gamma}}^T \mathbf{I} \hat{\boldsymbol{\gamma}} \\ &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{s}_j(\boldsymbol{\beta}) \right) \mathbf{I}^{-1} \mathbf{I} \left( \mathbf{I}^{-1} \sum_{t_j} \mathbf{G}_j \mathbf{s}_j(\boldsymbol{\beta}) \right) \\ &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{s}_j(\boldsymbol{\beta}) \right)^T \mathbf{I}^{-1} \left( \sum_{t_j} \mathbf{G}_j \mathbf{s}_j(\boldsymbol{\beta}) \right) \\ &= \left( \sum_{t_j} \mathbf{G}_j \mathbf{s}_j(\boldsymbol{\beta}) \right)^T \left( \sum_{t_j} \mathbf{G}_j \mathbf{V}_j \mathbf{G}_j \right)^{-1} \left( \sum_{t_j} \mathbf{G}_j \mathbf{s}_j(\boldsymbol{\beta}) \right),\end{aligned}\tag{3.22}$$

which corresponds to the score test (2.27), and under the null hypothesis is asymptotic  $\chi^2$ -distributed with  $p$  degrees of freedom.

### 3.3.2 Case II: When $\boldsymbol{\beta}$ is unknown

Now assume that  $\boldsymbol{\beta}$  is unknown and let  $\hat{\boldsymbol{\beta}}$  be the maximum partial likelihood estimate under the null hypothesis of  $\boldsymbol{\gamma} = \mathbf{0}$  where  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)^T$ . Further, let  $\hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}) = \mathbf{x}_{i_j} - \mathbf{E}(\hat{\boldsymbol{\beta}}, t_j)$  be the Schoenfeld residuals and  $\hat{\mathbf{V}}_j = \mathbf{V}(\hat{\boldsymbol{\beta}}, t_j)$  for ease of the notation. Then by (3.14) the scaled Schoenfeld residuals takes the form of

$$\hat{\mathbf{s}}_j^*(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{V}}_j^{-1} \hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}), \quad (3.23)$$

and by (3.15) the expectation is given by

$$\mathbb{E}[\hat{\mathbf{s}}_j^*(\hat{\boldsymbol{\beta}})] = \mathbb{E}[\hat{\mathbf{V}}_j^{-1} \hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}})] \approx \mathbf{G}_j \boldsymbol{\gamma}. \quad (3.24)$$

Since  $\hat{\boldsymbol{\beta}}$  is the value of  $\boldsymbol{\beta}$  such that  $\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$  follows from (2.21), we have  $\sum_{t_j} \hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}) = \mathbf{0}$  which implies that the Schoenfeld residuals are correlated. The covariance matrix follows from Grambsch and Therneau (1994) which is expressed as

$$\text{var}[\hat{\mathbf{s}}_j^*(\hat{\boldsymbol{\beta}})] = \text{var}[\hat{\mathbf{V}}_j^{-1} \hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}})] \approx \hat{\mathbf{V}}_j^{-1} - \left( \sum_{t_k} \hat{\mathbf{V}}_k \right)^{-1}. \quad (3.25)$$

An estimate of  $\boldsymbol{\gamma}$  may be obtained by using the multivariate generalized least squares, which gives

$$\tilde{\boldsymbol{\gamma}} = \mathbf{D}(\hat{\boldsymbol{\beta}})^{-1} \sum_{t_j} \mathbf{G}_j \hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}), \quad (3.26)$$

where

$$\mathbf{D}(\hat{\boldsymbol{\beta}}) = \sum_{t_j} \mathbf{G}_j \hat{\mathbf{V}}_j \mathbf{G}_j - \left( \sum_{t_j} \mathbf{G}_j \hat{\mathbf{V}}_j \right) \left( \sum_{t_j} \hat{\mathbf{V}}_j \right)^{-1} \left( \sum_{t_j} \mathbf{G}_j \hat{\mathbf{V}}_j \right)^T. \quad (3.27)$$

The matrix  $\mathbf{D}(\hat{\boldsymbol{\beta}})$  is found by solving the  $\mathbf{I}_{jk}$ 's from the partitioned matrix (2.29) using that  $\mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{t_j} \mathbf{V}(\boldsymbol{\beta}, \boldsymbol{\gamma}, t_j)$  and  $\mathbf{S}^{(i)}(\boldsymbol{\beta}, \boldsymbol{\gamma}, t_j)$  given in (3.19), and then using (2.31) to find the submatrix  $\mathbf{I}^{22}$  under the null hypothesis of  $\boldsymbol{\gamma} = \mathbf{0}$  and the maximum partial likelihood estimator  $\hat{\boldsymbol{\beta}}$ . The inverse of the submatrix  $\mathbf{D}(\hat{\boldsymbol{\beta}})^{-1}$  gives a consistent estimator of the covariance matrix of  $\tilde{\boldsymbol{\gamma}}$  under the null hypothesis. Moreover, a test-statistic under the null hypothesis becomes

$$\begin{aligned} \chi_{\tilde{\boldsymbol{\gamma}}}^2 &= \tilde{\boldsymbol{\gamma}}^T \mathbf{D}(\hat{\boldsymbol{\beta}}) \tilde{\boldsymbol{\gamma}} \\ &= \left( \sum_{t_j} \mathbf{G}_j \hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}) \right)^T \mathbf{D}(\hat{\boldsymbol{\beta}})^{-1} \left( \sum_{t_j} \mathbf{G}_j \hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}}) \right), \end{aligned} \quad (3.28)$$

which corresponds to the score test (2.33) based on the maximum partial likelihood  $\hat{\boldsymbol{\beta}}$ . The test-statistic (3.28) is asymptotic  $\chi^2$ -distributed with  $p$  degrees of freedom when the proportional hazards assumption holds.

### 3.3.3 Approximation of the score test when $\boldsymbol{\beta}$ is unknown

The weighted covariance matrix  $\hat{\mathbf{V}}_j = \mathbf{V}(\hat{\boldsymbol{\beta}}, t_j)$  at time  $t_j$  for the Schoenfeld residuals,  $\hat{\mathbf{s}}_j(\hat{\boldsymbol{\beta}})$ , may be unstable in practice. Because when it is near the end of the study it may happen that very few individuals are left such that the number of individuals in the risk



set is less than the number of rows of  $\widehat{\mathbf{V}}_j$ . In this case,  $\widehat{\mathbf{V}}_j$  will be singular. The variation of  $\widehat{\mathbf{V}}_j$  is slowly for most data sets, and is quite stable until the last few events of interest occurs. If we combine this observation with the fact that

$$\sum_{t_j} \widehat{\mathbf{V}}_j = \mathbf{I}(\widehat{\boldsymbol{\beta}}), \quad (3.29)$$

where  $\mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1}$  is the covariance matrix of  $\widehat{\boldsymbol{\beta}}$ , it suggests the use of the approximation

$$\widehat{\mathbf{V}}_j \approx \bar{\mathbf{V}} = \mathbf{I}(\widehat{\boldsymbol{\beta}})d^{-1}, \quad (3.30)$$

where  $d$  is the number of uncensored event times. Now redefine the diagonal elements for the  $p \times p$  diagonal matrix  $\mathbf{G}_j$  to be  $g(t_j) - \bar{g}$  (note that this will only ease the further calculation without changing the estimate of  $\gamma$ ), where  $\bar{g}$  is the mean of the  $g(t_j)$ 's. For ease of the notation, we let  $g(t_j) = g_j$ . Obviously, we have that

$$\sum_{t_j} \mathbf{G}_j = \sum_{t_j} (g_j - \bar{g}) = 0. \quad (3.31)$$

By plugging (3.30) into (3.27), we get

$$\begin{aligned} \mathbf{D}(\widehat{\boldsymbol{\beta}}) &\approx \sum_{t_j} \mathbf{G}_j \bar{\mathbf{V}} \mathbf{G}_j - \left( \sum_{t_j} \mathbf{G}_j \bar{\mathbf{V}} \right) \left( d \bar{\mathbf{V}} \right)^{-1} \left( \sum_{t_j} \mathbf{G}_j \bar{\mathbf{V}} \right)^T \\ &= \sum_{t_j} \mathbf{G}_j [\mathbf{I}(\widehat{\boldsymbol{\beta}})d^{-1}] \mathbf{G}_j \\ &= d^{-1} \sum_{t_j} \mathbf{G}_j \mathbf{I}(\widehat{\boldsymbol{\beta}}) \mathbf{G}_j, \end{aligned} \quad (3.32)$$

where the second term in the first equality is zero by (3.31). Then by using column-row expansion of the  $p \times p$  matrices  $\mathbf{G}_j$  and  $\mathbf{I}(\widehat{\boldsymbol{\beta}})$  in the third equality of (3.32), we get

$$\mathbf{D}(\widehat{\boldsymbol{\beta}}) \approx d^{-1} \sum_{t_j} (g_j - \bar{g})^2 \mathbf{I}(\widehat{\boldsymbol{\beta}}). \quad (3.33)$$

Now, define  $\widehat{\mathbf{S}}$  to be the  $d \times p$  matrix of Schoenfeld residuals where the rows are the transposed of the  $p \times 1$  vectors  $\widehat{\mathbf{s}}_1(\widehat{\boldsymbol{\beta}}), \widehat{\mathbf{s}}_2(\widehat{\boldsymbol{\beta}}), \dots, \widehat{\mathbf{s}}_d(\widehat{\boldsymbol{\beta}})$ . Also similarly define the  $d \times p$  matrix of scaled Schoenfeld residuals  $\widehat{\mathbf{S}}^*$ . Then by transposing (3.23), we have that

$$[\widehat{\mathbf{s}}_j^*(\widehat{\boldsymbol{\beta}})]^T = [\widehat{\mathbf{s}}_j(\widehat{\boldsymbol{\beta}})]^T \widehat{\mathbf{V}}_j^{-1}. \quad (3.34)$$

This gives the scaled Schoenfeld residuals matrix to be

$$\widehat{\mathbf{S}}^* \approx \widehat{\mathbf{S}} \bar{\mathbf{V}}^{-1} = d \widehat{\mathbf{S}} \mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1}. \quad (3.35)$$

Further, we can find that

$$\begin{aligned} \sum_{t_j} \mathbf{G}_j \widehat{\mathbf{s}}_j(\widehat{\boldsymbol{\beta}}) &= (g_1 - \bar{g}) \widehat{\mathbf{s}}_1(\widehat{\boldsymbol{\beta}}) + (g_2 - \bar{g}) \widehat{\mathbf{s}}_2(\widehat{\boldsymbol{\beta}}) + \dots + (g_d - \bar{g}) \widehat{\mathbf{s}}_d(\widehat{\boldsymbol{\beta}}) \\ &= \begin{bmatrix} \widehat{\mathbf{s}}_1(\widehat{\boldsymbol{\beta}}) \\ \widehat{\mathbf{s}}_2(\widehat{\boldsymbol{\beta}}) \\ \vdots \\ \widehat{\mathbf{s}}_d(\widehat{\boldsymbol{\beta}}) \end{bmatrix}^T \times (\mathbf{g} - \bar{g}) \\ &= \widehat{\mathbf{S}}^T (\mathbf{g} - \bar{g}) \end{aligned} \quad (3.36)$$

where  $\mathbf{g} = (g_1, g_2, \dots, g_d)^T$  such that  $(\mathbf{g} - \bar{\mathbf{g}}) = (g_1 - \bar{g}, g_2 - \bar{g}, \dots, g_d - \bar{g})^T$ . Hence, by plugging (3.33), (3.35) and (3.36) into the test-statistic (3.28), we get the global test of proportional hazards assumption over all covariates to be

$$\begin{aligned}
T &= \left( \hat{\mathbf{S}}^T (\mathbf{g} - \bar{\mathbf{g}}) \right)^T \left( d^{-1} \sum_{t_j} (g_j - \bar{g})^2 \mathbf{I}(\hat{\boldsymbol{\beta}}) \right)^{-1} \left( \hat{\mathbf{S}}^T (\mathbf{g} - \bar{\mathbf{g}}) \right) \\
&= \frac{(\mathbf{g} - \bar{\mathbf{g}})^T \hat{\mathbf{S}} \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1} \hat{\mathbf{S}}^T (\mathbf{g} - \bar{\mathbf{g}})}{d^{-1} \sum_{t_j} (g_j - \bar{g})^2} \\
&= \frac{(\mathbf{g} - \bar{\mathbf{g}})^T d^{-1} \hat{\mathbf{S}}^* \mathbf{I}(\hat{\boldsymbol{\beta}}) \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1} d^{-1} \mathbf{I}(\hat{\boldsymbol{\beta}}) \hat{\mathbf{S}}^{*T} (\mathbf{g} - \bar{\mathbf{g}})}{d^{-1} \sum_{t_j} (g_j - \bar{g})^2} \\
&= \frac{(\mathbf{g} - \bar{\mathbf{g}})^T \hat{\mathbf{S}}^* \mathbf{I}(\hat{\boldsymbol{\beta}}) \hat{\mathbf{S}}^{*T} (\mathbf{g} - \bar{\mathbf{g}})}{d \sum_{t_j} (g_j - \bar{g})^2},
\end{aligned} \tag{3.37}$$

and under the null hypothesis is approximately  $\chi^2$ -distributed with  $p$  degrees of freedom.

We may also find a test-statistic for the univariate test of proportionality assumption for the  $k$ th covariate. First of all, note that  $\hat{\mathbf{S}} \approx d^{-1} \hat{\mathbf{S}}^* \mathbf{I}(\hat{\boldsymbol{\beta}})$ , which is just a rewritten form of (3.35). By inserting this into (3.36), and further inserting both (3.33) and (3.36) into the estimator (3.26), one obtain that

$$\tilde{\gamma} \approx \frac{\hat{\mathbf{S}}^{*T} (\mathbf{g} - \bar{\mathbf{g}})}{\sum_{t_j} (g_j - \bar{g})^2}.$$

The univariate test of proportionality for the  $k$ th covariate is based on

$$\tilde{\gamma}_k \approx \frac{\sum_{t_j} (g_j - \bar{g}) \hat{\mathbf{s}}_{jk}^*(\hat{\boldsymbol{\beta}})}{\sum_{t_j} (g_j - \bar{g})^2}, \tag{3.38}$$

where  $\hat{\mathbf{s}}_{jk}^*(\hat{\boldsymbol{\beta}})$  is the corresponding scaled Schoenfeld residuals for the  $k$ th covariate. As mentioned earlier, the matrix  $\mathbf{D}(\hat{\boldsymbol{\beta}})^{-1}$  gives a consistent estimator of the covariance matrix of  $\tilde{\gamma}$  under the null hypothesis. Hence, we estimate the variance of  $\tilde{\gamma}_k$  by

$$D^{kk}(\hat{\boldsymbol{\beta}}) = \frac{d}{\sum_{t_j} (g_j - \bar{g})^2} I^{kk}(\hat{\boldsymbol{\beta}}), \tag{3.39}$$

where  $I^{ij}(\hat{\boldsymbol{\beta}}) = \mathbf{I}_{ij}(\hat{\boldsymbol{\beta}})^{-1}$  is the  $(i, j)$ -element of the covariance matrix  $\mathbf{I}(\hat{\boldsymbol{\beta}})^{-1}$  of  $\hat{\boldsymbol{\beta}}$ . A test-statistic for the univariate test of proportionality assumption for the  $k$ th covariate becomes

$$T_k = \frac{\tilde{\gamma}_k^2}{D^{kk}(\hat{\boldsymbol{\beta}})} = \frac{\left( \sum_{t_j} (g_j - \bar{g}) \hat{\mathbf{s}}_{jk}^*(\hat{\boldsymbol{\beta}}) \right)^2}{I^{kk}(\hat{\boldsymbol{\beta}}) \sum_{t_j} (g_j - \bar{g})^2} d^{-1}, \tag{3.40}$$

which is approximately  $\chi^2$ -distributed with one degree of freedom when the proportionality assumption holds.

### Computation of the tests by using R

In the statistical software package **R** (R Development Core Team, 2014), the use of the `cox.zph` command with the argument `transform="log"` corresponds to the test of the proportional hazards assumption for a Cox regression model fit (`coxph`). The argument `transform="log"` is a character string specifying how the survival times should be transformed before the test is performed. The Kaplan-Meier time-transform function, that is  $g(t) = \hat{S}(t)$ , is default and will be applied when no argument is specified in the `cox.zph` command. Other types of time-transform functions are also available, but we will concentrate on  $\hat{S}(t)$  and  $\log t$ . If everything is done correctly, the command window in **R** should automatically print out a summary of the tests. The column `chisq` from the output is the test-statistics based on (3.40), while the last row `GLOBAL` gives the global test based on (3.37). The column `p` gives the  $p$ -values for the corresponding tests.

Note also that the `cox.zph` command with the `transform="log"` argument is not the same test as the one in Subsection 3.2.1 using the `tt()` function with  $g(t) = \log t$ . The difference here is that the model (3.8) includes all  $p$  time-dependent terms in the model formula, and then use the test-statistic (3.40) for checking the proportional hazards assumption for the  $k$ th covariate. On the other hand, the model (3.4) just includes the  $q$  first time-dependent terms in the model formula. Thus these two models are not directly comparable.

#### 3.3.4 Illustration: The German Breast Cancer Study Data

As an illustration, the German Breast Cancer Study Data will be used to produce results based on the test statistics (3.37) and (3.40). We will use both type of time-transform functions as mentioned in the computational procedure above such that a comparison is possible to be performed. A summary of both tests for all covariates are given in Table 3.3.

A comparison of the  $p$ -values shows that both the univariate test of proportionality gives the same conclusion. That is, both the tests shows that the covariate Tumor Grade 3 has a non-proportional effect. This conclusion is also consistent with what we have found by using the methods in Subsection 3.2.1 and Subsection 3.2.2. Further, the global test of proportionality assumption does not holds in both cases, and the reason is that the non-proportional effect of the covariate Tumor Grade 3. In addition, it is also worth

**Table 3.3** Results by using  $g(t) = \log t$  and  $g(t) = \hat{S}(t)$  from the univariate test of proportionality  $T_k$  given in (3.40) and the global test  $T$  given in (3.37) of the proportional hazards assumption over all covariates.

Covariate	$g(t) = \log t$		$g(t) = \hat{S}(t)$	
	$\chi^2$ -values	$\Pr(>\chi^2)$	$\chi^2$ -values	$\Pr(>\chi^2)$
$\log_2(\text{No. of Nodes})$	1.598	0.206	0.919	0.338
$\log_2(\text{No. of Prog. R.})$	0.959	0.327	1.533	0.216
Hormone 2	0.060	0.807	0.021	0.884
Tumor Grade 2	1.211	0.271	1.148	0.284
Tumor Grade 3	4.672	0.031	4.216	0.040
GLOBAL	13.026	0.023	12.215	0.032

mentioning that the calculated  $p$ -values in both tests for the Number of Positive Lymph Nodes and the Number of Progesterone Receptor are quite different.

### 3.4 Tests based on martingale residuals

The following section is based on the book by Aalen et al. (2008) and the paper by Lin et al. (1993). The martingale residuals are an important and useful tool for checking the fit of Cox's regression model. It is the difference between the observed and expected numbers of events for each individual over the full study time interval. For the Cox's model, different plots and goodness-of-fit tests have been proposed based on these residuals. In the following section, we will consider the tests based on the score process and the  $\chi^2$ -test based on grouped martingale residual processes.

#### 3.4.1 The martingale residual processes

We want to define the martingale residual processes, but first of all we introduce the process

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(u) du, \quad (3.41)$$

where  $N_i(t)$  is the counting process defined in (2.10). Since  $dN_i(t)$  is a binary variable, we may rewrite (2.11) as

$$\lambda_i(t) dt = E[dN_i(t) \mid \text{past}]. \quad (3.42)$$

A reformulation of this relation gives

$$E[dN_i(t) - \lambda_i(t) dt \mid \text{past}] = 0, \quad (3.43)$$

where  $\lambda_i(t) dt$  can be moved inside the conditional expectation since it is a function of the past. Then we have that

$$E[dM_i(t) \mid \text{past}] = 0 \quad (3.44)$$

which shows that (3.41) is a martingale. Now introduce the cumulative intensity processes for the  $i$ th individual at time  $t$  to be

$$\begin{aligned} \Lambda_i(t) &= \int_0^t \lambda_i(u) du \\ &= \int_0^t Y_i(u) \alpha(t | \mathbf{x}_i) du \\ &= \int_0^t Y_i(u) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\} \alpha_0(u) du; \quad i = 1, \dots, n, \end{aligned} \quad (3.45)$$

where we used the intensity process  $\lambda_i(t)$  given in (2.12) with  $\alpha(t | \mathbf{x}_i)$  given in (2.17). If we in (3.45) replace  $\boldsymbol{\beta}$  with  $\hat{\boldsymbol{\beta}}$  and  $\alpha_0(u) du$  with the increment  $d\hat{A}_0(u)$  given in (2.25), we get the estimated cumulative intensity processes

$$\hat{\Lambda}_i(t) = \int_0^t Y_i(u) \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i\} d\hat{A}_0(u) = \sum_{t_j \leq t} \frac{Y_i(t_j) \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i\}}{\sum_{l \in R_j} \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{x}_l\}}. \quad (3.46)$$

Thus by using (3.41) the martingale residual processes is given by

$$\widehat{M}_i(t) = N_i(t) - \widehat{\Lambda}_i(t); \quad i = 1, \dots, n, \quad (3.47)$$

where  $N_i(t)$  and  $\widehat{\Lambda}_i(t)$  are respectively the observed and expected numbers of events for the  $i$ th individual at time  $t$ . If we let  $\tau$  be the upper time limit for the study, then we get the martingale residuals

$$\widehat{M}_i = \widehat{M}_i(\tau) = N_i(\tau) - \widehat{\Lambda}_i(\tau). \quad (3.48)$$

### 3.4.2 Tests based on the score process

For survival data, the usefulness of each of the martingale residual processes is not much since they contains too little information. However, useful plots and goodness-of-fit tests may be obtained by aggregating them over all or groups of individuals. Thus we may consider the process

$$U(\widehat{\boldsymbol{\beta}}, t) = \sum_{i=1}^n \mathbf{x}_i \widehat{M}_i(t). \quad (3.49)$$

If we insert  $\widehat{M}_i(t)$  as given in (3.47) into (3.49), we get

$$\begin{aligned} U(\widehat{\boldsymbol{\beta}}, t) &= \sum_{i=1}^n \mathbf{x}_i \{N_i(t) - \widehat{\Lambda}_i(t)\} \\ &= \sum_{i=1}^n \mathbf{x}_i \left\{ \int_0^t dN_i(u) - \int_0^t Y_i(u) \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_i\} d\widehat{A}_0(u) \right\} \\ &= \sum_{i=1}^n \left\{ \int_0^t \mathbf{x}_i dN_i(u) - \int_0^t \mathbf{x}_i \frac{Y_i(u) \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_i\}}{\sum_{l=1}^n Y_l(u) \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_l\}} dN_{\bullet}(u) \right\} \\ &= \sum_{i=1}^n \left\{ \int_0^t \mathbf{x}_i dN_i(u) - \int_0^t \frac{\sum_{l=1}^n Y_l(u) \mathbf{x}_l \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_l\}}{\sum_{l=1}^n Y_l(u) \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_l\}} dN_i(u) \right\} \\ &= \sum_{i=1}^n \int_0^t \left\{ \mathbf{x}_i - \frac{S^{(1)}(\widehat{\boldsymbol{\beta}}, u)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, u)} \right\} dN_i(u) \\ &= \sum_{i=1}^n \int_0^t \left\{ \mathbf{x}_i - \mathbf{E}(\widehat{\boldsymbol{\beta}}, u) \right\} dN_i(u), \end{aligned} \quad (3.50)$$

where  $d\widehat{A}_0(u)$  is the differential of (2.25), while  $\mathbf{E}(\widehat{\boldsymbol{\beta}}, u)$  is given in (3.10). If we evaluate (3.50) at  $t = \tau$ , we obtain the score equations (2.21) with  $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ . For this reason  $U(\widehat{\boldsymbol{\beta}}, t)$  is called the score process. The score process (3.50) may also be written in the following form

$$U(\widehat{\boldsymbol{\beta}}, t) = \sum_{t_j \leq t} \left\{ \mathbf{x}_{i_j} - \mathbf{E}(\widehat{\boldsymbol{\beta}}, t_j) \right\}, \quad (3.51)$$

which is a sum of the Schoenfeld residuals. For the  $k$ th covariate, the score process is given by

$$U_k(\widehat{\boldsymbol{\beta}}, t) = \sum_{i=1}^n \int_0^t \left\{ x_{ik} - \mathbf{E}_k(\widehat{\boldsymbol{\beta}}, u) \right\} dN_i(u), \quad (3.52)$$

where  $E_k(\hat{\boldsymbol{\beta}}, u)$  is the weighted mean for the  $k$ th covariate at time  $u$ . We may also rewrite (3.52) in the form of

$$U_k(\hat{\boldsymbol{\beta}}, t) = \sum_{t_j \leq t} \{x_{ijk} - E_k(\hat{\boldsymbol{\beta}}, t_j)\}, \quad (3.53)$$

which is a sum of the Schoenfeld residuals for the  $k$ th covariate.

According to the paper by Lin et al. (1993), a test of proportionality based on the score process is given by

- Unweighted test statistic:

$$T_{w_0} = \sup_t |U_k(\hat{\boldsymbol{\beta}}, t)| \quad (3.54)$$

- Weighted test statistic:

$$T_{w_1} = \sup_t \sum_{k=1}^p [I^{kk}(\hat{\boldsymbol{\beta}})]^{1/2} |U_k(\hat{\boldsymbol{\beta}}, t)|, \quad (3.55)$$

where  $I^{ij}(\hat{\boldsymbol{\beta}}) = I_{ij}(\hat{\boldsymbol{\beta}})^{-1}$  is the  $(i, j)$ -element of the covariance matrix  $I(\hat{\boldsymbol{\beta}})^{-1}$  of  $\hat{\boldsymbol{\beta}}$ .

When checking the proportionality assumption for the  $k$ th covariate, one can plot  $U_k(\hat{\boldsymbol{\beta}}, t)$  vs. time  $t$ . However, interpretation of this plot is not easy. The problem is that the distribution to the score process (3.52) is difficult to find. For this reason, Lin et al. (1993) suggest the use of simulation such that one is able to approximate the distribution of the observed score process. Then by making plots of the observed and a number of simulated score processes one can conduct graphical inspections. Note that when  $t = \tau$  is the upper time limit for the study, then the score process (3.52) is zero. If the proportionality assumption holds, one can expect that the plot of the observed score process corresponds to the  $k$ th covariates are fluctuating around zero on the time (horizontal) axis. In addition, the observed score process should not be far away from the simulated one.

The  $p$ -value for the test statistics (3.54) and (3.55) is determined by simulation of the score processes. For the unweighted (weighted) test statistic, the  $p$ -value is equivalent to the proportion of all simulated (weighted) score processes that has a supremum of the absolute value larger than the supremum of the corresponding (weighted) observed score process measured in absolute value.

### Computation of the tests by using R

To check the proportional hazards assumption based on the test statistics (3.54) and (3.55) in **R** (R Development Core Team, 2014), one have to use the `cox.aalen` function from the `timereg` package (instead of the `coxph` function as mentioned earlier) to fit the Cox regression model. For every covariates in the model formula, apply the argument `prop(covariate)` for proportional effect. In addition, one should also give the following arguments in the `cox.aalen` function: Use either `weighted.test=0` for unweighted test or `weighted.test=1` for weighted test, `n.sim=1000` for 1000 simulations in the re-sampling of the score process, and the options `rate.sim=0` and `residuals=1` are used to obtain residuals that can be used for model validation. Then by using the `summary` command, one gets the results of the fit and the test of proportionality for all covariates in the model formula. If one apply the `plot` command on the fitted model with the additional

argument `specific.comps=c(.,.)` (e.g. `c(1,2)` for the first and second covariate), then **R** will produce plots of the observed and simulated score processes for the corresponding covariate.

### 3.4.3 Illustration: The German Breast Cancer Study Data

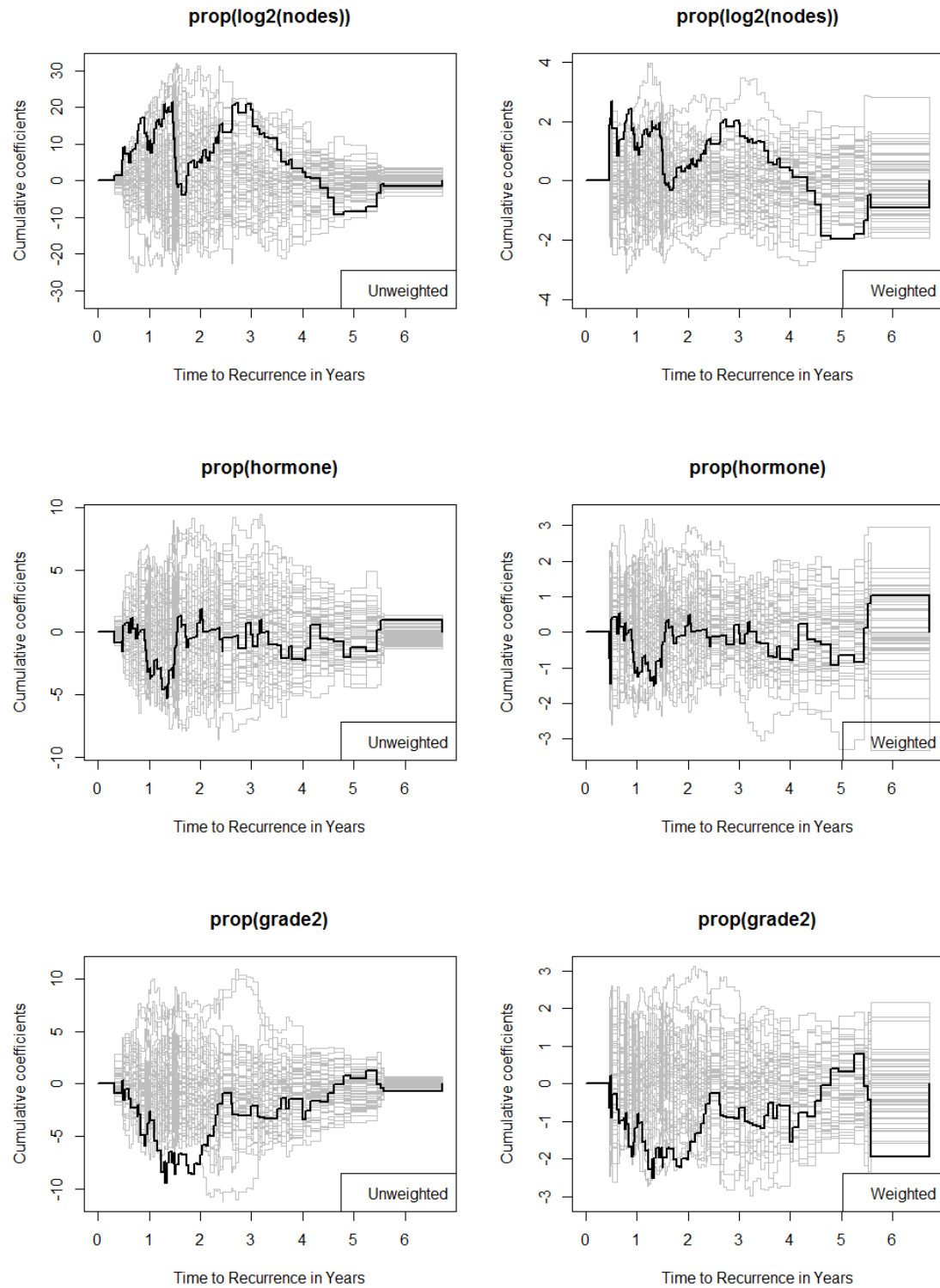
As an illustration, we will apply the computation procedure as mentioned above by using the German Breast Cancer Study Data. The results based on 1000 simulations in the re-sampling of the score process are shown in Table 3.4. It follows from the  $p$ -values that the log-transformed Number of Positive Lymph Nodes and Tumor Grade 3 are significant for the test of proportionality in both of the tests.

The plots of the observed and simulated score processes for each covariates in the fitted model are shown in Figure 3.1 and Figure 3.2. Note that the “dark line” corresponding to the observed score process, while those “gray colored curves” are the simulated score processes. Further, note also that those plots on the left hand-side represent the unweighted score process, while the right hand-side is the weighted score process. As we mentioned earlier, if the proportionality holds, the observed score process of each covariates should fluctuate around zero on the time axis. By comparing the curves for the observed and the simulated score process for the log-transformed Number of Positive Lymph Nodes, the similarity seems quite good for both unweighted and weighted score process. For Hormone Therapy in both cases, the observed score process is fluctuating around zero on the time axis, which also satisfy with the  $p$ -values from Table 3.4. For Tumor Grade 2, the observed (unweighted) score process between the 1st and 2nd years seems to lie under the simulated one. In this case, the fit of the Cox model is not so impressive.

In Figure 3.2, we have the log-transformed Progesterone Receptors and the Tumor Grade 3 which is significant for both of the tests of proportionality. By looking at the plot corresponding to the Progesterone Receptors in both cases, the observed score process is going negative during the period between the 1st and 2nd years compared with the simulated one, which indicate that the fit of the Cox model is poor. Further, the plot of the observed score process for the Tumor Grade 3 in both cases seems to be significantly different from the simulated one during the period between the 1st and 3rd years. The calculated  $p$ -values shows also that both tests of proportionality are significant with respectively 0.001 and 0.003. This conclusion is also consistent with what we have found by using the other methods shown earlier.

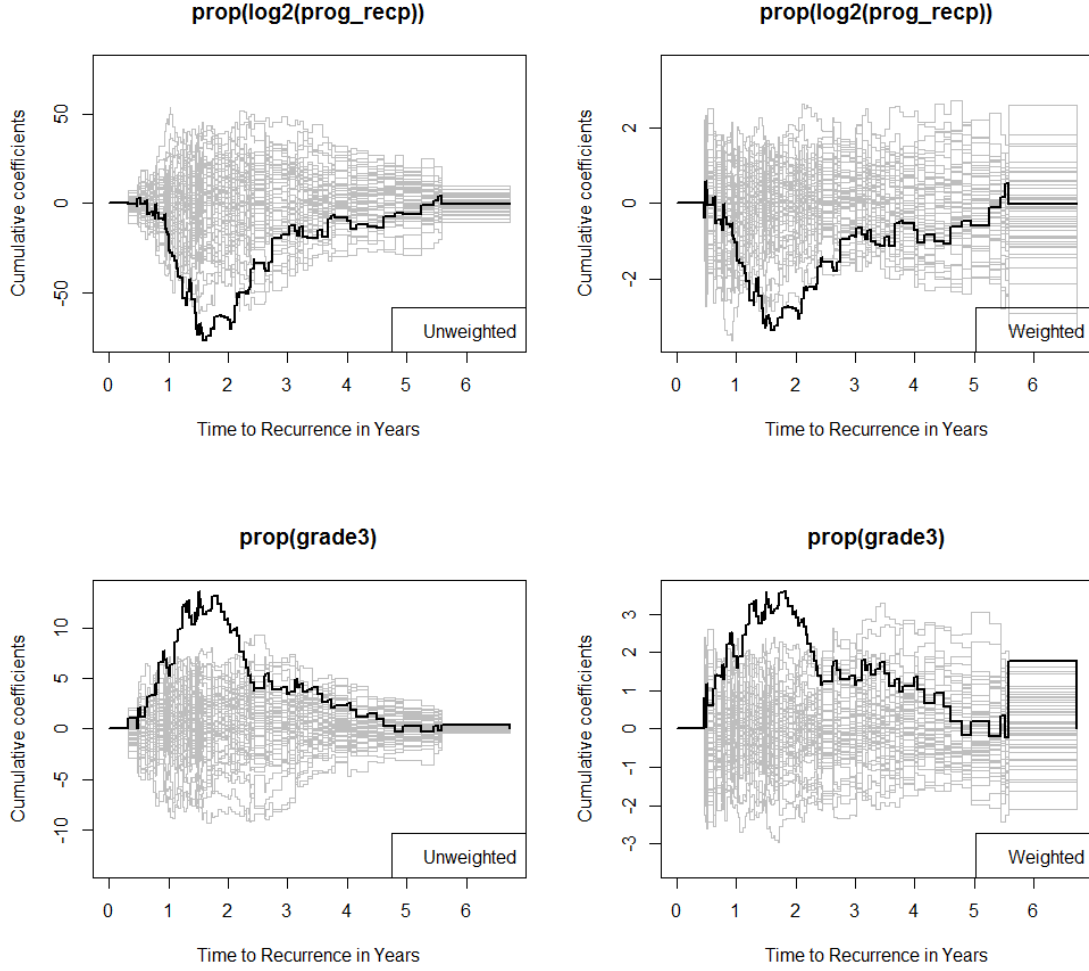
**Table 3.4** Results from the tests of proportional hazards assumption using the test statistics (3.54) and (3.55) based on 1000 simulations in the re-sampling of the score process.

Covariates	Unweighted test		Weighted test	
	$T_{w_0}$	$\Pr(>T_{w_0})$	$T_{w_1}$	$\Pr(>T_{w_1})$
$\log_2(\text{No. of Nodes})$	21.6	0.295	2.7	0.139
$\log_2(\text{No. of Prog. R.})$	76.6	0.005	3.7	0.017
Hormone 2	5.3	0.673	1.5	0.858
Tumor Grade 2	9.5	0.076	2.5	0.195
Tumor Grade 3	13.6	0.001	3.6	0.003



**Figure 3.1** Plots of the score processes for the log-transformed Number of Positive Lymph Nodes, Hormone Therapy and Tumor Grade 2. Left hand-side panel: Unweighted score processes. Right hand-side panel: Weighted score processes.





**Figure 3.2** Plots of the score processes for the log-transformed Progesterone Receptors and Tumor Grade 3. Left hand-side panel: Unweighted score processes. Right hand-side panel: Weighted score processes.

#### 3.4.4 The $\chi^2$ -test based on grouped martingale residual processes

As mentioned at the beginning in the previous subsection, we may obtain useful plots and goodness-of-fit tests by aggregating the martingale residual processes over all or groups of individuals. Assume that  $J, I = 1, 2, \dots, G$  denote groups of individuals, typically based on the values of one or two covariates. Let  $J(u)$  and  $I(u)$  be the sets of all individuals who belongs to group  $J$  and  $I$  at time  $u$ , respectively. Then the grouped martingale residual process for group  $J$  becomes

$$\widehat{M}_J(t) = \int_0^t \sum_{i \in J(u)} d\widehat{M}_i(u) = N_J(t) - \sum_{t_j \leq t} \frac{\sum_{i \in R_j \cap J(t_j)} \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_i\}}{\sum_{l \in R_j} \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_l\}}, \quad (3.56)$$

while for group  $I$

$$\widehat{M}_I(t) = \int_0^t \sum_{i \in I(u)} d\widehat{M}_i(u) = N_I(t) - \sum_{t_j \leq t} \frac{\sum_{i \in R_j \cap I(t_j)} \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_i\}}{\sum_{l \in R_j} \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_l\}}, \quad (3.57)$$

where  $N_J(t) = \int_0^t \sum_{i \in J(u)} dN_i(u)$  and  $N_I(t) = \int_0^t \sum_{i \in I(u)} dN_i(u)$  are the observed number of events in group  $J$  and  $I$  in the interval  $[0, t]$ , respectively. The last term on the right-hand side of (3.56) and (3.57) is an estimate of the expected number of events in the group when the model (2.17) holds true. Since we have to estimate the regression coefficients, the grouped martingale residual processes are only approximately martingales. It is possible to show that the grouped martingale residual processes in large samples are approximately normally distributed with mean zero when the model (2.17) holds true.

Let  $s$  and  $t$  be the time such that  $s \leq t$ . Then the covariance between  $\widehat{M}_I(s)$  and  $\widehat{M}_J(t)$  may be estimated by

$$\widehat{\sigma}_{IJ}(s, t) = \widehat{\phi}_{IJ}(0, s) - \widehat{\boldsymbol{\Psi}}_I(0, s)^T \mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\Psi}}_J(0, t), \quad (3.58)$$

where  $\mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1}$  is the covariance matrix of  $\widehat{\boldsymbol{\beta}}$ . Here

$$\widehat{\phi}_{IJ}(u_1, u_2) = \sum_{u_1 < t_j \leq u_2} \frac{S_I^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)} \left\{ \delta_{IJ} - \frac{S_J^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)} \right\} \quad (3.59)$$

and

$$\widehat{\boldsymbol{\Psi}}_J(u_1, u_2) = \sum_{u_1 < t_j \leq u_2} \left\{ \frac{S_J^{(1)}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)} - \frac{S_J^{(0)}(\widehat{\boldsymbol{\beta}}, t_j) S^{(1)}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)^2} \right\}, \quad (3.60)$$

where  $\delta_{IJ}$  is 1 when  $I = J$  and 0 otherwise. Further,  $S^{(0)}(\boldsymbol{\beta}, t_j)$  and  $S^{(1)}(\boldsymbol{\beta}, t_j)$  are defined in (2.36) and (2.37), respectively. The definition of  $S_J^{(0)}(\boldsymbol{\beta}, t_j)$  and  $S_J^{(1)}(\boldsymbol{\beta}, t_j)$  are similar to  $S^{(0)}(\boldsymbol{\beta}, t_j)$  and  $S^{(1)}(\boldsymbol{\beta}, t_j)$ , but with the sums restricted to the individuals who belong to group  $J$  at time  $t$ . The corresponding definition for group  $I$  is similar.

Useful plots may be provided in the sense that one use the martingale residual processes to derive formal goodness-of-fit tests. Formally, one may use a  $\chi^2$ -test based on a comparison of the observed and expected number of events in the  $G$  groups in  $K$  disjoint time intervals. Let  $H, L = 1, 2, \dots, K$  be the disjoint time intervals and  $0 = \tau_0 < \tau_1 < \dots < \tau_{K-1} < \tau_K = \tau$  be a partitioning of the study time interval. Now introduce

$$\widehat{M}_{HJ} = \widehat{M}_J(\tau_H) - \widehat{M}_J(\tau_{H-1}) = O_{HJ} - E_{HJ}, \quad (3.61)$$

where  $O_{HJ} = N_J(\tau_H) - N_J(\tau_{H-1})$  is the observed number of events in group  $J$  in time interval  $H$ , while

$$E_{HJ} = \sum_{\tau_{H-1} < t_j \leq \tau_H} \frac{\sum_{i \in R_j \cap J(t_j)} \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_i\}}{\sum_{l \in R_j} \exp\{\widehat{\boldsymbol{\beta}}^T \mathbf{x}_l\}} \quad (3.62)$$

is the corresponding expected number under model (2.17). The definition is similar for group  $I$  in time interval  $L$ . Note that the sum of (3.56) and (3.57) is zero at any given time  $t$ . Therefore we disregard the contribution from, example, the first group when deriving a  $\chi^2$  goodness-of-fit test. Then consider the  $K(G-1)$ -vector  $\widehat{\mathbf{M}}$  with elements  $\widehat{M}_{HJ}$  for

$H = 1, 2, \dots, K$  and  $J = 2, 3, \dots, G$ . By the results of the grouped martingale residual processes, it follows that the vector  $\widehat{\mathbf{M}}$  in large samples is approximately multivariate normally distributed with mean zero when model (2.17) holds true. Its covariance matrix may be estimated by

$$\widehat{\mathbf{\Sigma}} = \{\widehat{\sigma}_{LI,HJ}\} \quad (3.63)$$

with elements

$$\begin{aligned} \widehat{\sigma}_{LI,HJ} &= \widehat{\text{Cov}}(\widehat{M}_{LI}, \widehat{M}_{HJ}) \\ &= \delta_{LH} \widehat{\phi}(\tau_{H-1}, \tau_H) - \widehat{\Psi}_I(\tau_{L-1}, \tau_L)^T \mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\Psi}_J(\tau_{H-1}, \tau_H); \end{aligned} \quad (3.64)$$

$H, L = 1, 2, \dots, K$ ;  $J, I = 1, 2, \dots, G$ ; where  $\delta_{LH}$  is 1 when  $L = H$  and 0 otherwise. Then a goodness-of-fit test is based on the statistic

$$\chi^2 = \widehat{\mathbf{M}}^T \widehat{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{M}}, \quad (3.65)$$

which in large samples is approximately  $\chi^2$ -distributed with  $K(G-1)$  degrees of freedom when model (2.17) holds true.

Consider the extension of model (2.17) where an individual  $i$  who belongs to group  $J$  at time  $t \in (\tau_{H-1}, \tau_H]$  has the following hazard rate

$$\alpha(t|\mathbf{x}_i) = \alpha_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i + \gamma_{HJ}\}. \quad (3.66)$$

Then a hypothesis of testing the proportional hazards assumption is equivalent to the test of additional  $K(G-1)$  parameters  $\gamma_{HJ}$  are all equal to zero. Thus the goodness-of-fit statistic  $\chi^2$  in (3.65) corresponds to the score test (2.33). The score test statistic in this case is approximately  $\chi^2$ -distributed with  $K(G-1)$  degrees of freedom when the model (2.17) holds true. In standard statistical software package, the  $\chi^2$  goodness-of-fit test can be computed as the score test for the addition of categorical grouping variables.

### 3.4.5 Illustration: The German Breast Cancer Study Data

We want to illustrate the use of the test statistics (3.65) by fitting a model of the form of (3.66). That is, we may use the same method as described in Subsection 3.2.2 to divide the time interval into groups with  $K = 3$  groups. Since the covariates Hormone Therapy and Tumor Grade are already categorized and tested in Subsection 3.2.2, we are therefore interested in the numeric covariates Number of Positive Lymph Nodes and Number of Progesterone Receptor in the following illustration. We may, e.g. divide the numeric covariates in three groups, so  $G = 3$ . Thus, we want to fit the extended Cox model of the form of

$$\alpha(t|\mathbf{x}_i) = \alpha_0(t) \exp \left\{ \boldsymbol{\beta}^T \mathbf{x}_i + \sum_{l=1}^K (\gamma_{l2} + \gamma_{l3}) \right\}, \quad (3.67)$$

where we have chosen  $\gamma_{l1}$  as the reference group in the  $l$ th time interval. The coefficients in the extended model (3.67) may be explained as follows:  $\gamma_{l2}$  denotes an individual  $i$  who belongs to the 2nd group at the  $l$ th time interval, while  $\gamma_{l3}$  denotes an individual  $i$  who belongs to the 3rd group at the  $l$ th time interval. A test of the proportional hazards assumption corresponds to the test of  $\gamma_{HJ} = 0$  for all  $H = 1, 2, 3$  and  $J = 2, 3$ . The results

**Table 3.5** Results from the test of proportional hazards assumption with one additional grouped covariate in the model (2.17) at a time.

<b>Grouped covariate</b>	$\chi_{LR}^2$	$\chi_{SC}^2$	$\chi_W^2$	$\Pr(>\chi_{SC,df=6}^2)$
Number of Positive Lymph Nodes	12.9	12.5	12.3	0.051
Number of Progesterone Receptors	10.0	9.4	9.2	0.152

of the tests are given in Table 3.5 which is based on one additional grouped covariate in the model (2.17) at a time. Hence, it follows from the  $p$ -values from the score test statistics that both additional grouped covariate are non-significant, which implies that the proportional hazards assumption holds.

## Chapter 4

# Model checking by simulations

In Chapter 3 we have given an overview of the methods and tests that may be used for checking the proportional hazards assumption for Cox's regression model. We have also used **R** (R Development Core Team, 2014) to illustrate how each of them are performed based on real data from the German Breast Cancer Study Data. In order to perform a thorough comparison of the performance of the tests under different circumstances, we will use simulation in the following chapter to generate our own data set of survival times for both cases of proportional and non-proportional hazards. In Section 4.1 we will define the general expression for the survival time. In Section 4.2 we will consider the general procedure for simulating a data set of survival times when the model is correctly specified followed by an illustration. In Section 4.3 we will look at two opposite cases, that is, when the model is incorrectly specified followed by two illustrations.

### 4.1 General considerations of survival times

Assume that the assumption of proportional hazards holds for the Cox regression model (2.17). Then by plugging (2.17) into (2.3), the cumulative hazard rate for the  $i$ th individual becomes

$$A(t|\mathbf{x}_i) = A_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}, \quad (4.1)$$

where  $A_0(t)$  is the cumulative baseline hazard function. Further, by plugging (4.1) into (2.4), the survival function for the  $i$ th individual takes the form of

$$S(t|\mathbf{x}_i) = \exp\left\{-A_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}\right\}. \quad (4.2)$$

The cumulative distribution function of Cox regression model (2.17) is given by

$$F(t|\mathbf{x}_i) = 1 - S(t|\mathbf{x}_i) = 1 - \exp\left\{-A_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}\right\}. \quad (4.3)$$

Let  $Y$  be a random variable with distribution function  $F$ . Then  $U = F(Y)$  is uniformly distributed on the interval from 0 to 1. Further, if  $U \sim U(0, 1)$ , then  $(1 - U) \sim U(0, 1)$ . We let  $T_i$  be the survival time for an individual  $i$  of the Cox regression model (2.17). Then it follows from (4.3) that

$$U_i = \exp\left\{-A_0(T_i) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i\}\right\} \sim U(0, 1). \quad (4.4)$$

If  $\alpha_0(t) > 0$  for all  $t$ , then the survival time  $T_i$  can be expressed as

$$T_i = A_0^{-1}\left(-\log(U_i) \exp\{-\boldsymbol{\beta}^T \mathbf{x}_i\}\right), \quad (4.5)$$

where  $A_0^{-1}(t)$  is the inverse of the cumulative baseline hazard function  $A_0(t)$ , and  $U_i$  is a random variable with  $U_i \sim U(0, 1)$ .

## 4.2 When the model is correctly specified

### 4.2.1 A general procedure for simulating a data set of survival times

For simulating a data set of survival times in **R**, one can make a user-written function in the following form

```
myfunction <- function(arg1, arg2, ... ){
  statements
  return(object)
}
```

In the “**statements**” part of the function, the following variables for the  $n$  individuals have to be defined:

- The variable **id** is a sequence of numbers that runs from 1 to  $n$ . This is used to identify the survival times for each of the individuals.
- The random variables  $U_1, U_2, \dots, U_n$  are independent and generated from the uniform distribution on the interval 0 to 1.
- The vector of survival times  $\mathbf{T} = (T_1, T_2, \dots, T_n)^T$ . In order to obtain an expression for the survival time  $T_i$ , we may choose the cumulative baseline hazard function to have a Weibull form, that is

$$A_0(t) = (1/b)^a t^a,$$

where  $a$  is the shape parameter and  $b$  is the scale parameter. It gives the inverse cumulative baseline hazard function to be

$$A_0^{-1}(t) = bt^{1/a}.$$

Then based on the expression (4.5) the survival time  $T_i$  for the  $i$ th individual is generated from

$$T_i = \left(-b^a \log(U_i) \exp\{-\boldsymbol{\beta}^T \mathbf{x}_i\}\right)^{1/a}; \quad \text{for } i = 1, 2, \dots, n.$$

- The vector of censoring times  $\mathbf{C} = (C_1, C_2, \dots, C_n)^T$  is generated from the exponential distribution with a censoring rate  $\lambda$ . The censoring rate  $\lambda$  should be chosen such that one gets a particular proportion of censored observations.

- The vector of observed survival times  $\tilde{\mathbf{T}} = (\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_n)^T$ . This is defined for the  $i$ th individual as

$$\tilde{T}_i = \min(T_i, C_i, \tau); \quad \text{for } i = 1, 2, \dots, n,$$

where  $\tau$  is the upper time limit for the study. In **R**, one can obtain the minimum value by using the `pmin` command.

- The vector of censoring indicators  $\mathbf{D} = (D_1, D_2, \dots, D_n)^T$ . This is already defined in (2.5) as  $D_i$  for an individual  $i$ . The censoring indicator  $D_i$  can be obtained in **R** by using the `as.numeric( $\tilde{T}_i == T_i$ )` command, which returns the value of 1 if the statement is true and 0 otherwise.

Introduce the notation  $\mathbf{x}_k^n = (x_{1k}, x_{2k}, \dots, x_{nk})^T$ , for  $k = 1, 2, \dots, p$ , for the vector of the values of the  $k$ th covariate for  $n$  individuals. Note that usually the vector of covariates for an individual  $i$  is given as  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ . Finally, by using the `data.frame(id,  $\mathbf{T}$ ,  $\mathbf{C}$ ,  $\tilde{\mathbf{T}}$ ,  $\mathbf{D}$ ,  $\mathbf{x}_1^n, \mathbf{x}_2^n, \dots, \mathbf{x}_p^n$ )` command in **R**, it will create a data set tightly coupled of all variables. This is the object that will be returned in the function above.

Moreover, the function should take the following parameters as the arguments for simulating the data set of survival times: The  $n$  number of individuals in the study, the  $p \times 1$  vector of coefficients  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ , the vector of covariates for the  $i$ th individual  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ , the shape parameter  $a$ , the scale parameter  $b$ , the censoring rate  $\lambda$ , and the parameter  $\tau$ .

#### 4.2.2 Check the proportionality of the correctly specified model

The following subsection is an illustration for simulating a data set of survival times when  $n = 250$  and  $n = 1000$  individuals. The simulated data set of survival times can be used to fit the Cox regression model, and by using the different methods of tests in Chapter 3 to check the assumption of proportionality. In this way, we can perform a thorough comparison of the performance of the tests when the model is correctly specified.

Let  $\tau = 1$  be the upper time limit for the study and the vector of covariates  $\mathbf{x}_i = (x_{i1}, x_{i2})^T$  to fit the Cox regression model (2.17). The first covariate  $x_{i1}$  is generated from the standard normal distribution, while the second covariate  $x_{i2}$  is generated from the Bernoulli distribution with success probability  $p = 0.5$ . Further, let  $\boldsymbol{\beta} = (0.5, 1)^T$  such that the Cox model for the  $i$ th individual takes the form of

$$\alpha(t|\mathbf{x}_i) = \alpha_0(t) \exp\{0.5x_{i1} + x_{i2}\}, \quad (4.6)$$

where the baseline hazard function  $\alpha_0(t) = (a/b^a)t^{a-1}$  takes the Weibull form. We want to keep the event rate at around 50%, and a reasonable choice is to let the shape parameter  $a = 0.5$ , the scale parameter  $b = 4.5$ , and the censoring rate  $\lambda = (1/b)^a \approx 0.471$ . These arguments will be applied in the function as mentioned in Subsection 4.2.1 based on 1000 simulations for both  $n = 250$  and  $n = 1000$  individuals. Further, the Cox model (4.6) will be fitted and checked for model misspecification in each simulation.

The achieved significance level, which is equivalent to the proportion of  $p$ -values  $\leq 5\%$  based on 1000 simulations, for  $x_1$  and  $x_2$  in each of the test is given in Table 4.1. Note that the different designations of the tests in this table have the following meaning:

- “T-D terms:  $\log t$ ” corresponds to (3.4) with  $g(t) = \log t$ .
- “T-D terms: Interval” corresponds to (3.5).
- “cox.zph with  $\log t$ ” corresponds to (3.40) with  $g(t) = \log t$ .
- “cox.zph with  $\hat{S}(t)$ ” corresponds to (3.40) with  $g(t) = \hat{S}(t)$  (the Kaplan-Meier time-transform function).
- “Unweighted score process” corresponds to (3.54).
- “Weighted score process” corresponds to (3.55).
- “cox.zph Global test:  $\log t$ ” corresponds to (3.37) with  $g(t) = \log t$ .
- “cox.zph Global test:  $\hat{S}(t)$ ” corresponds to (3.37) with  $g(t) = \hat{S}(t)$ .

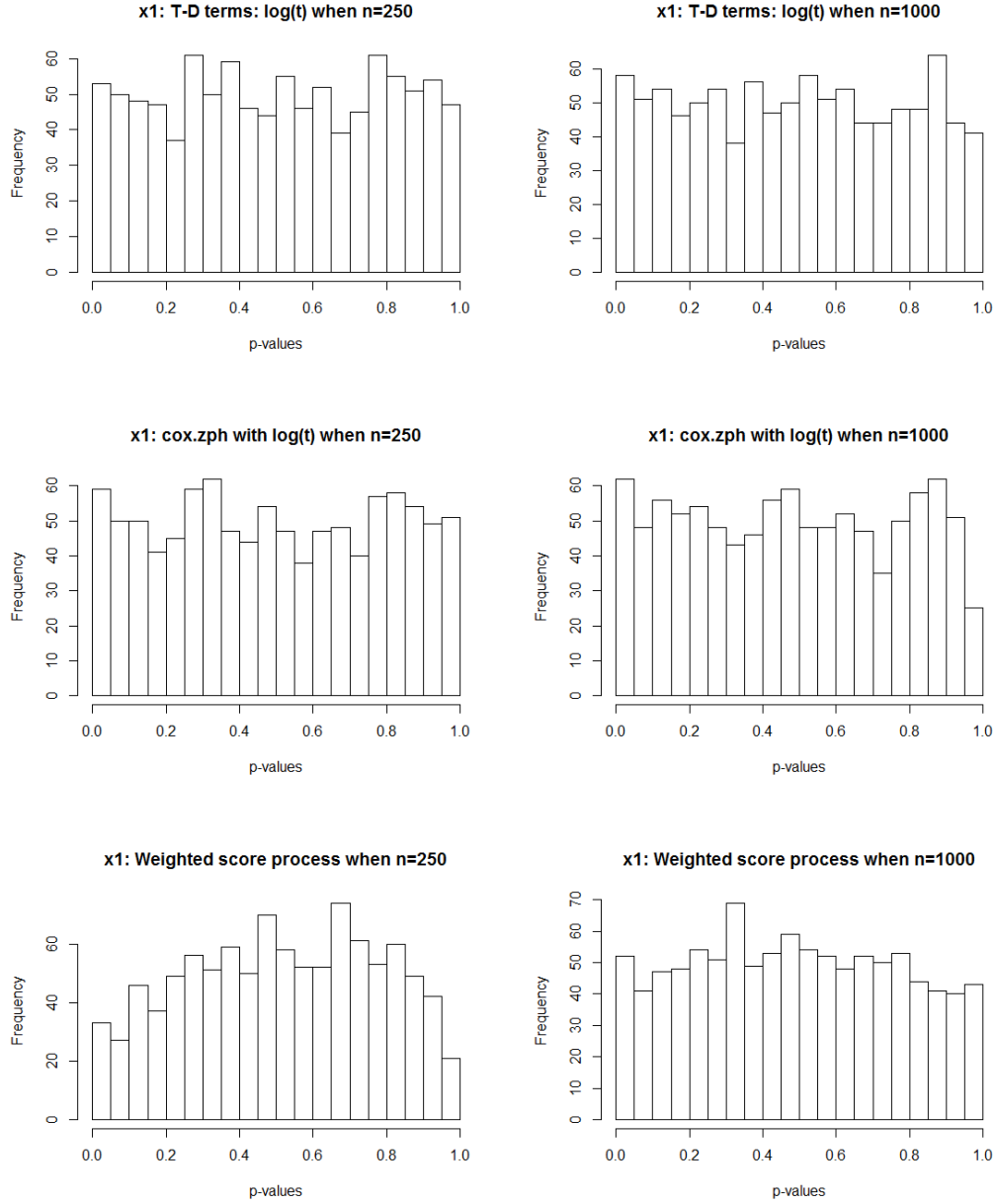
According to Table 4.1, we see that the achieved significance level for  $x_1$  and  $x_2$  for both groups of individuals in those tests are close to the nominal 5% level. This shows that the different tests achieves the correct level in most of the cases over the simulations. Note, however, that with  $n = 250$  the test with the “Weighted score process” for  $x_1$  and  $x_2$  does not obtain the nominal 5% level, the achieved level being 3.3% and 7.8%, respectively. When  $n = 1000$ , the same test shows that the number of non-proportional cases are improved for  $x_1$  with 5.2%, but still a bit large for  $x_2$  with 7.4%. Thus, we may conclude that the “Weighted score process” in this context is not satisfactory compared with the other tests.

Histograms of  $p$ -values from the test “T-D terms:  $\log t$ ”, “cox.zph with  $\log t$ ” and “Weighted score process” for  $x_1$  and  $x_2$  are shown in Figure 4.1 and Figure 4.2, respectively. Note that the left panels corresponding to the case with  $n = 250$  individuals, while the right panels corresponding to  $n = 1000$  individuals. The other histograms of  $p$ -values for the other tests are moved to Appendix A. The point here is that when the proportionality assumption holds, the  $p$ -values should be uniformly distributed. This is the case for all the tests for both  $n = 250$  and  $n = 1000$ , except for the test “Weighted score process”. For instance, the “Weighted score process” for  $x_1$  in the left-hand side of Figure 4.1 shows that

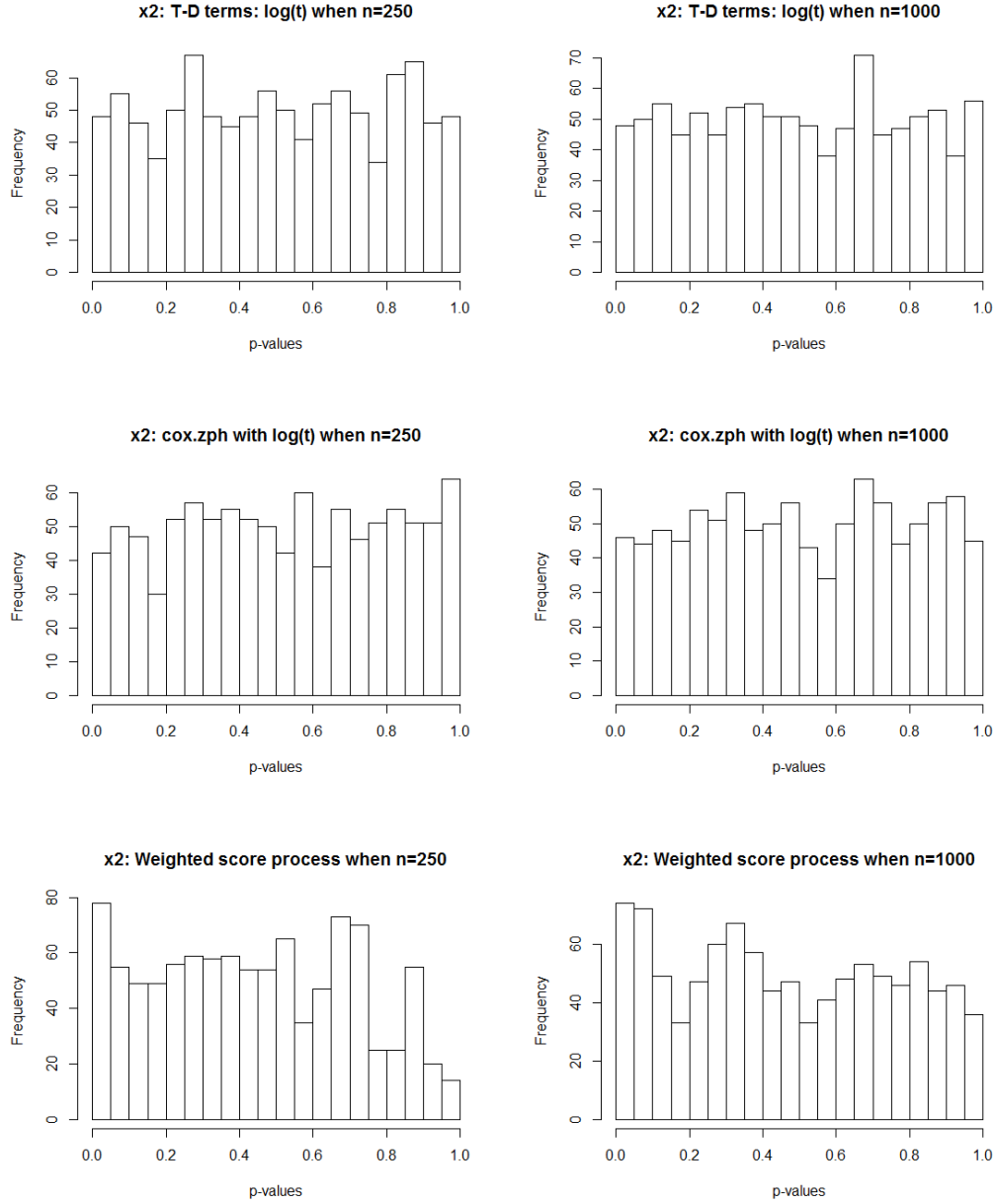
**Table 4.1** Achieved significance level for the test of proportionality for  $x_1$  and  $x_2$  over 1000 simulated data sets of survival times for both  $n = 250$  and  $n = 1000$  individuals. The Cox regression model (4.6) is correctly specified with  $\boldsymbol{\beta} = (0.5, 1)^T$ .

Test	$n = 250$		$n = 1000$	
	$x_1$	$x_2$	$x_1$	$x_2$
T-D terms: $\log t$	5.3%	4.8%	5.8%	4.8%
T-D terms: Interval	4.8%	5.1%	4.9%	4.7%
cox.zph with $\log t$	5.9%	4.2%	6.2%	4.6%
cox.zph with $\hat{S}(t)$	4.0%	4.5%	5.6%	5.9%
Unweighted score process	5.0%	5.0%	6.6%	5.1%
Weighted score process	3.3%	7.8%	5.2%	7.4%
cox.zph Global test: $\log t$	5.2%		5.3%	
cox.zph Global test: $\hat{S}(t)$	5.2%		6.1%	





**Figure 4.1** Histograms of  $p$ -values for  $x_1$  over 1000 simulated data sets of survival times for  $n = 250$  individuals (left panels) and  $n = 1000$  individuals (right panels) based on three different tests of proportionality when the model is correctly specified with  $\boldsymbol{\beta} = (0.5, 1)^T$ .



**Figure 4.2** Histograms of  $p$ -values for  $x_2$  over 1000 simulated data sets of survival times for  $n = 250$  individuals (left panels) and  $n = 1000$  individuals (right panels) based on three different tests of proportionality when the model is correctly specified with  $\boldsymbol{\beta} = (0.5, 1)^T$ .

too few  $p$ -values in the interval of  $[0, 0.1]$  are observed. On the other hand, the histograms for the same test for  $x_2$  in Figure 4.2 shows that too many small  $p$ -values are observed.

A problem worth mentioning is when fitting the model (3.4) with  $n = 1000$  individuals by using the `tt()` function in **R**. The estimated time for one simulation is approximate 75 seconds. Thus by using 1000 simulations, the laptop will need around 20-21 hours.

### 4.3 When the model is incorrectly specified

#### 4.3.1 Model with time-varying coefficients

If one or more coefficients in a Cox model is time-varying, then the assumption of proportional hazards is violated. We want to illustrate this situation by changing the model (4.6) such that one of the coefficients is time-varying. From now on, the hazard rate for the  $i$ th individual is given by

$$\alpha(t|x_i) = \alpha_0(t) \exp\{\beta_1 x_{i1} + \beta_2(t) x_{i2}\}, \quad (4.7)$$

where  $\alpha_0(t)$  is the baseline hazard with the Weibull form, that is  $\alpha_0(t) = (a/b^a)t^{a-1}$ , and  $\beta_1$  is a coefficient. Same as before, we will let  $x_{i1} \sim N(0, 1)$  and  $x_{i2} \sim \text{Bernoulli}(0.5)$ . The only difference is  $\beta_2(t)$  which is no longer a constant, but a function of  $t$ . We will let this function to be given as

$$\beta_2(t) = \beta_2 + g(t), \quad (4.8)$$

where  $\beta_2$  is a coefficient and  $g(t)$  is a known function.

#### When the covariate $x_{i2} = 0$

Since  $x_{i2}$  is Bernoulli distributed, the hazard rate for an individual  $i$  when  $x_{i2} = 0$  is given by

$$\alpha(t|x_{i1}) = (a/b^a)t^{a-1} \exp\{\beta_1 x_{i1}\}.$$

The corresponding cumulative hazard rate is

$$A(t|x_{i1}) = (1/b^a)t^a \exp\{\beta_1 x_{i1}\}.$$

Similarly, based on (4.5) the corresponding survival time  $T_i$  for an individual  $i$  is given by

$$T_i = \left( -b^a \log(U_i) \exp\{-\beta_1 x_{i1}\} \right)^{1/a}. \quad (4.9)$$

#### When the covariate $x_{i2} = 1$

When  $x_{i2} = 1$ , the hazard rate for an individual  $i$  is given by

$$\alpha(t|x_{i1}) = (a/b^a)t^{a-1} \exp\{\beta_1 x_{i1} + \beta_2 + g(t)\}. \quad (4.10)$$

The corresponding cumulative hazard rate takes the form of

$$A(t|x_{i1}) = (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2\} \int_0^t u^{a-1} \exp\{g(u)\} du. \quad (4.11)$$

The question now is which choice of the function  $g(t)$  and the constant  $a$  in (4.11) will make it possible to find an explicit expression for the inverse of the cumulative hazard rate. A possible choice of  $g(t)$  is

$$g(t) = \log(\gamma_0 + \gamma_1 t) \quad (4.12)$$

where  $\gamma_0 > 0$ . If  $\gamma_1 < 0$ , we will have  $\gamma_0 + \gamma_1 t < 0$  when  $t > -\gamma_0/\gamma_1$ . Thus for  $t$  large enough, the function (4.12) is not defined. Consequently when  $\gamma_1 < 0$ , a reformulation of the function (4.12) has to be given. That is,

$$g(t) = \begin{cases} \log(\gamma_0 + \gamma_1 t), & \text{if } t < -\frac{\gamma_0}{\gamma_1} \\ -\infty, & \text{if } t \geq -\frac{\gamma_0}{\gamma_1}. \end{cases} \quad (4.13)$$

The choice of  $\gamma_0$  and  $\gamma_1$  are given in the following way:

- When  $t = 0$  such that  $g(0) = c$ , we obtain:

$$\log(\gamma_0 + \gamma_1 \times 0) = c \quad \implies \quad \gamma_0 = e^c \quad (4.14)$$

where  $c \in \mathbb{R}$ .

- When  $t = t_0$  such that  $g(t_0) = 0$ , we obtain:

$$\log(\gamma_0 + \gamma_1 t_0) = 0 \quad \implies \quad \gamma_1 = \frac{1 - \gamma_0}{t_0} = \frac{1 - e^c}{t_0} \quad (4.15)$$

where  $t_0 > 0$ .

Assume first that  $c < 0$  such that  $\gamma_0 \in (0, 1)$  and  $\gamma_1 = \frac{1 - \gamma_0}{t_0} > 0$ . By inserting (4.12) into (4.11), we get

$$\begin{aligned} A(t|x_{i1}) &= (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2\} \int_0^t u^{a-1} (\gamma_0 + \gamma_1 u) du \\ &= (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2\} \left( \frac{\gamma_0}{a} t^a + \frac{\gamma_1}{a+1} t^{a+1} \right). \end{aligned}$$

In further calculations, we will let  $a = 1$  such that it is possible to find an explicit expression for the inverse of the cumulative hazard rate. Then we have that

$$A(t|x_{i1}) = (1/b) \exp\{\beta_1 x_{i1} + \beta_2\} (\gamma_0 t + \frac{1}{2} \gamma_1 t^2). \quad (4.16)$$

Further, let  $T_i$  be the survival time for an individual  $i$ , and the corresponding random variable  $U_i$  be given as

$$U_i = \exp\{-A(T_i|x_{i1})\} \sim U(0, 1).$$

This gives

$$\begin{aligned} -\log(U_i) &= A(T_i|x_{i1}) \\ &= (1/b) \exp\{\beta_1 x_{i1} + \beta_2\} \left( \gamma_0 T_i + \frac{1}{2} \gamma_1 T_i^2 \right). \end{aligned}$$

A reformulation of this equation gives the quadratic equation

$$\frac{1}{2}\gamma_1 T_i^2 + \gamma_0 T_i + b \log(U_i) \exp\{-\beta_1 x_{i1} - \beta_2\} = 0,$$

with the solutions of the survival time  $T_i$  to be either

$$T_i = \frac{-\gamma_0 + \sqrt{\gamma_0^2 - 2b\gamma_1 \log(U_i) \exp\{-\beta_1 x_{i1} - \beta_2\}}}{\gamma_1} \quad (4.17)$$

or

$$T_i = \frac{-\gamma_0 - \sqrt{\gamma_0^2 - 2b\gamma_1 \log(U_i) \exp\{-\beta_1 x_{i1} - \beta_2\}}}{\gamma_1}. \quad (4.18)$$

Since  $\gamma_1 > 0$  and  $\log(U_i) < 0$ , it is obvious that the value inside the square root in (4.17) is positive, such that the survival time  $T_i$  is positive or zero if and only if the following inequality holds

$$-\gamma_0 + \sqrt{\gamma_0^2 - 2b\gamma_1 \log(U_i) \exp\{-\beta_1 x_{i1} - \beta_2\}} \geq 0 \quad \Leftrightarrow \quad U_i \leq 1. \quad (4.19)$$

Since  $U_i \sim U(0, 1)$ , the inequality (4.19) will always holds. Similarly, it is clear that the numerator in (4.18) is negative when  $\gamma_1 > 0$  such that the survival times becomes negative. Hence for  $\gamma_1 > 0$ , the survival time  $T_i$  is generated from (4.17). Note that (4.18) will not be considered in further calculations.

Now, assume that  $c > 0$  such that  $\gamma_0 \in (1, \infty)$  and  $\gamma_1 = \frac{1-\gamma_0}{t_0} < 0$ . Then, by inserting (4.13) into (4.11), a reformulation of the cumulative hazard rate is given as follows:

- If  $t < -\frac{\gamma_0}{\gamma_1}$ , then by inserting the first equation in (4.13) into (4.11), we obtain

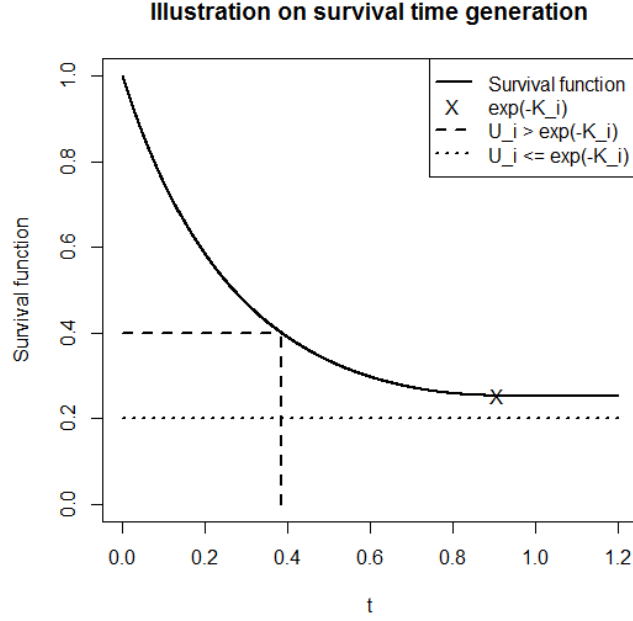
$$A(t|x_{i1}) = (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2\} \left( \frac{\gamma_0}{a} t^a + \frac{\gamma_1}{a+1} t^{a+1} \right).$$

- If  $t \geq -\frac{\gamma_0}{\gamma_1}$ , then integrating from 0 to  $-\frac{\gamma_0}{\gamma_1}$  and from  $-\frac{\gamma_0}{\gamma_1}$  to  $t$  by using (4.13) and (4.11), we obtain

$$\begin{aligned} A(t|x_{i1}) &= (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2\} \left[ \int_0^{-\gamma_0/\gamma_1} u^{a-1} (\gamma_0 + \gamma_1 u) du \right. \\ &\quad \left. + \int_{-\gamma_0/\gamma_1}^t u^{a-1} \exp\{-\infty\} du \right] \\ &= (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2\} \left[ \frac{\gamma_0}{a} \left( -\frac{\gamma_0}{\gamma_1} \right)^a + \frac{\gamma_1}{a+1} \left( -\frac{\gamma_0}{\gamma_1} \right)^{a+1} \right]. \end{aligned}$$

The corresponding simplified expression for  $a = 1$  is given by

$$A(t|x_{i1}) = \begin{cases} \left(\frac{1}{b}\right) \exp\{\beta_1 x_{i1} + \beta_2\} (\gamma_0 t + \frac{1}{2} \gamma_1 t^2), & \text{if } t < -\frac{\gamma_0}{\gamma_1} \\ -\left(\frac{1}{2b}\right) \left(\frac{\gamma_0^2}{\gamma_1}\right) \exp\{\beta_1 x_{i1} + \beta_2\}, & \text{if } t \geq -\frac{\gamma_0}{\gamma_1}. \end{cases} \quad (4.20)$$



**Figure 4.3** For a given value of  $x_{i1}$ , this is an illustration on how the survival time  $T_i$  for an individual  $i$  is generated when  $x_{i2} = 1$  and  $\gamma_1 < 0$ . For instance, if  $U_i = 0.4$ , then the corresponding survival time  $T_i = 0.384$ . On the other hand, if  $U_i = 0.2$ , then  $T_i = \infty$ . Note that  $K_i$  corresponds to the maximum value of the cumulative hazard rate for the  $i$ th individual such that  $\exp(-K_i)$  is the minimum of the survival function.

If  $t = -\frac{\gamma_0}{\gamma_1}$ , we obtain the maximum value of the cumulative hazard rate, and this is given by

$$\begin{aligned}
 K_i &= \max_t \{A(t|x_{i1})\} \\
 &= A\left(-\frac{\gamma_0}{\gamma_1}|x_{i1}\right) \\
 &= -\left(\frac{1}{2b}\right)\left(\frac{\gamma_0^2}{\gamma_1}\right) \exp\{\beta_1 x_{i1} + \beta_2\}.
 \end{aligned} \tag{4.21}$$

As illustrated in Figure 4.3, to generate the survival time  $T_i$ , we first generate  $U_i$  from the  $U(0, 1)$ -distribution, and then obtain  $T_i$  as follows:

- If  $U_i > \exp\{-K_i\}$ , then by using the first equation in (4.20), we obtain

$$T_i = \frac{-\gamma_0 + \sqrt{\gamma_0^2 - 2b\gamma_1 \log(U_i) \exp\{-\beta_1 x_{i1} - \beta_2\}}}{\gamma_1}. \tag{4.22}$$

- If  $U_i \leq \exp\{-K_i\}$ , then

$$T_i = \infty. \tag{4.23}$$

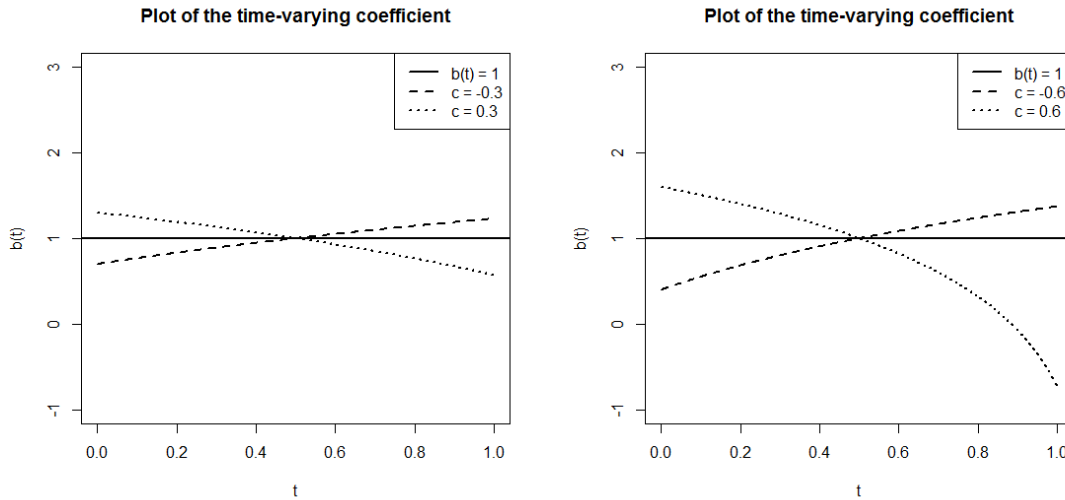
To sum up, the survival time  $T_i$  for the  $i$ th individual is generated from (4.9) when  $x_{i2} = 0$ . When  $x_{i2} = 1$  and “ $c < 0 \Rightarrow \gamma_1 > 0$ ”, then  $T_i$  is generated from (4.17). When  $x_{i2} = 1$  and “ $c > 0 \Rightarrow \gamma_1 < 0$ ”, then  $T_i$  is generated from (4.22) if  $U_i > \exp\{-K_i\}$  holds, else  $T_i = \infty$ .

### 4.3.2 Check the proportionality of the incorrectly specified model

To simulate our own data set of survival times, we will use the general procedure as described in Subsection 4.2.1 with  $T_i$  as shown in (4.9), (4.17) and (4.22). Since the constant  $c$  can be both positive and negative, we will have two non-proportional Cox model to be checked. The survival times will be based on both  $n = 250$  and  $n = 1000$  individuals in the study over 1000 simulations. The simulated data set of survival times will then be used to fit the Cox regression model. Further the different methods of tests in Chapter 3 will be applied to check the proportional hazards assumption.

Let  $\beta_1 = 0.5$ ,  $\beta_2 = 1$ ,  $\tau = 1$ , the shape parameter  $a = 1$  and the scale parameter  $b = 2$ . The covariate  $x_{i1}$  and  $x_{i2}$  are generated from  $N(0,1)$  and  $\text{Bernoulli}(0.5)$ , respectively. Further, it follows from (4.14) and (4.15) that  $g(0) = c$  for  $t = 0$  and  $g(t_0) = 0$  for  $t = t_0$ , respectively. Therefore, at time  $t = 0$  the function  $g(t)$  start from the point  $c$  and cross the point 0 at time  $t = t_0$ . In other words, the steepness of the curve for  $g(t)$  is given by the constant  $c$ . Note that at time  $t = t_0$  the time-varying coefficient  $\beta_2(t) = \beta_2$ . Thus, how the constant  $c$  and  $t_0$  should be chosen can be shown by making various plots of  $\beta_2(t)$  vs.  $t \in [0, 1]$ . A possible choice is  $t_0 = 0.5$  and  $c = \pm 0.3$ . The plot is shown in the left-hand side of Figure 4.4. Note that the horizontal line corresponding to the assumption of proportionality, that is,  $g(t) = 0$  for all  $t \in [0, 1]$ . Further, the plot on the right-hand side of Figure 4.4 corresponding to the case with  $c = \pm 0.6$ . This choice of  $c$  give us a curve which is twice steeper than the previous one, and corresponding to a stronger non-proportional effect of the incorrectly specified model.

Note that the censoring time  $C_i$ , censored survival time  $\tilde{T}_i$  and censoring indicator  $D_i$  are generated in the same way as mentioned in Subsection 4.2.1. Further, the event rate will still be kept at around 50%, and a reasonable choice of the censoring rate for  $c = -0.3$  and  $c = 0.3$  are  $\lambda = 0.315$  and  $\lambda = 0.370$ , respectively. For  $c = -0.6$  and  $c = 0.6$ , we have



**Figure 4.4** A plot of the time-varying coefficient  $\beta_2(t)$  vs. time  $t \in [0, 1]$ . Left panel: When  $c = \pm 0.3$ . Right panel: When  $c = \pm 0.6$ . The curves cross each other at time  $t = 0.5$ . The horizontal line corresponds to the assumption of proportionality.

**Table 4.2** Achieved significance level and achieved power for  $x_1$  and  $x_2$ , respectively, in the case of  $c = -0.3$  and  $c = -0.6$  over 1000 simulated data sets of survival times for both  $n = 250$  and  $n = 1000$  individuals. The Cox regression model (4.7) is incorrectly specified with  $\beta = (0.5, 1)^T$ .

Test	For $c = -0.3$				For $c = -0.6$			
	$n = 250$		$n = 1000$		$n = 250$		$n = 1000$	
	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$
T-D terms: $\log t$	5.2%	9.9%	4.6%	28.8%	4.7%	27.5%	6.4%	76.9%
T-D terms: Interval	4.6%	8.1%	4.9%	25.4%	4.9%	20.3%	6.5%	71.3%
cox.zph with $\log t$	5.8%	8.9%	4.3%	27.3%	4.1%	25.8%	5.4%	74.7%
cox.zph with $\hat{S}(t)$	4.4%	10.6%	4.7%	34.3%	3.7%	30.4%	5.5%	84.1%
Unweighted score process	5.5%	10.0%	5.0%	29.1%	5.1%	24.7%	6.0%	77.1%
Weighted score process	3.3%	8.2%	3.9%	22.6%	2.5%	17.6%	4.1%	67.3%
cox.zph Global test: $\log t$	7.5%		21.5%		21.2%		65.7%	
cox.zph Global test: $\hat{S}(t)$	9.4%		27.0%		20.9%		75.5%	

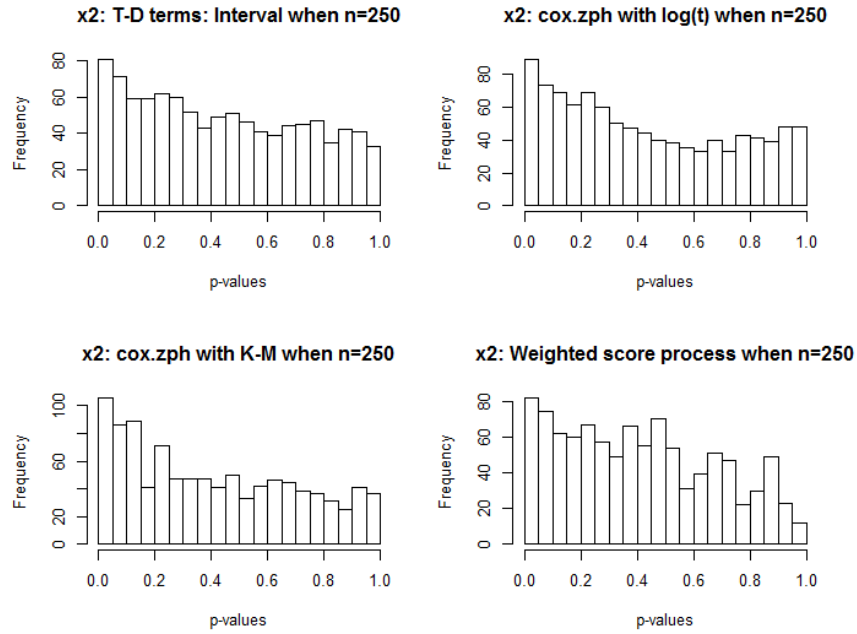
**Table 4.3** Achieved significance level and achieved power for  $x_1$  and  $x_2$ , respectively, in the case of  $c = 0.3$  and  $c = 0.6$  over 1000 simulated data sets of survival times for both  $n = 250$  and  $n = 1000$  individuals. The Cox regression model (4.7) is incorrectly specified with  $\beta = (0.5, 1)^T$ .

Test	For $c = 0.3$				For $c = 0.6$			
	$n = 250$		$n = 1000$		$n = 250$		$n = 1000$	
	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$
T-D terms: $\log t$	5.2%	11.3%	4.7%	35.0%	6.3%	33.4%	6.7%	89.0%
T-D terms: Interval	6.0%	10.0%	4.6%	31.4%	6.7%	30.8%	8.3%	88.3%
cox.zph with $\log t$	6.1%	10.5%	5.2%	31.7%	6.1%	30.0%	5.9%	85.7%
cox.zph with $\hat{S}(t)$	6.1%	14.9%	4.7%	46.1%	5.0%	44.5%	5.8%	96.3%
Unweighted score process	5.9%	15.0%	5.4%	41.7%	5.3%	46.4%	8.2%	96.0%
Weighted score process	3.0%	17.3%	4.4%	41.6%	3.9%	44.2%	7.8%	95.4%
cox.zph Global test: $\log t$	8.4%		24.8%		23.6%		78.0%	
cox.zph Global test: $\hat{S}(t)$	12.8%		36.7%		35.6%		94.0%	

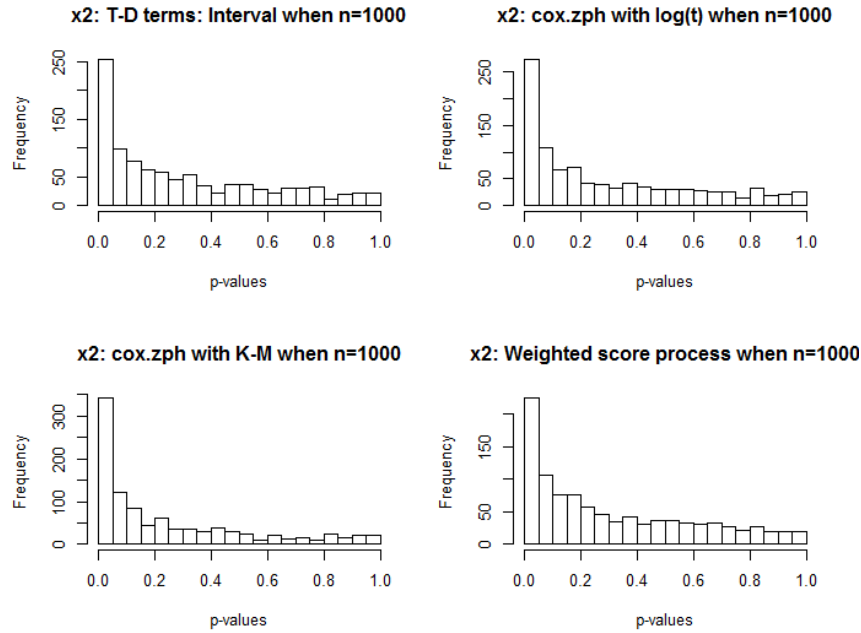
to change the rate to  $\lambda = 0.300$  and  $\lambda = 0.410$ , respectively.

Since  $x_1$  has a proportional effect, we want to obtain an achieved significance level close to the nominal 5% level. It follows from Table 4.2 for  $x_1$  in both cases with  $c = -0.3$  and  $c = -0.6$  that the test with the “Weighted score process” in three of the four cases does not achieve the correct level with respectively, 3.3%, 3.9% and 2.5%. On the other hand, since  $\beta_2(t)$  is a time-varying coefficient, we expect that the achieved power for  $x_2$  should be large for almost all tests of proportionality. Further, it follows from Table 4.2 that the test “cox.zph with  $\hat{S}(t)$ ” obtained the most optimal value of the achieved power for  $x_2$  in all cases, and consequently the most efficient test of non-proportionality compared with the other ones. At the same time, the test with the weakest results for both the achieved significance level and the achieved power for  $x_1$  and  $x_2$ , respectively, is the “Weighted score process”. Histograms of  $p$ -values for  $x_2$  from the test “T-D terms: Interval”, “cox.zph with  $\log t$ ”, “cox.zph with  $\hat{S}(t)$ ” and “Weighted score process” illustrating the results with  $c = -0.3$  are given in Figure 4.5 and Figure 4.6 for  $n = 250$  and  $n = 1000$  individuals,

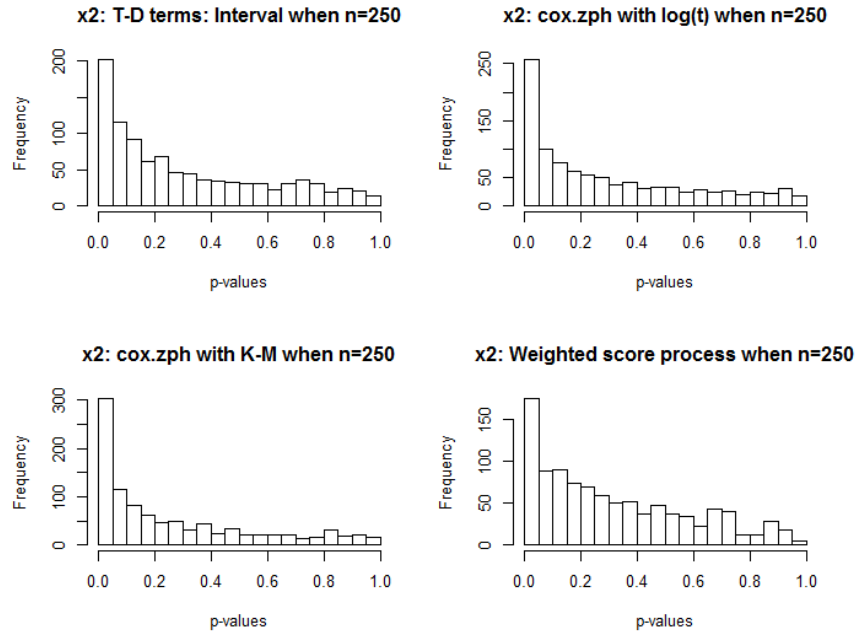




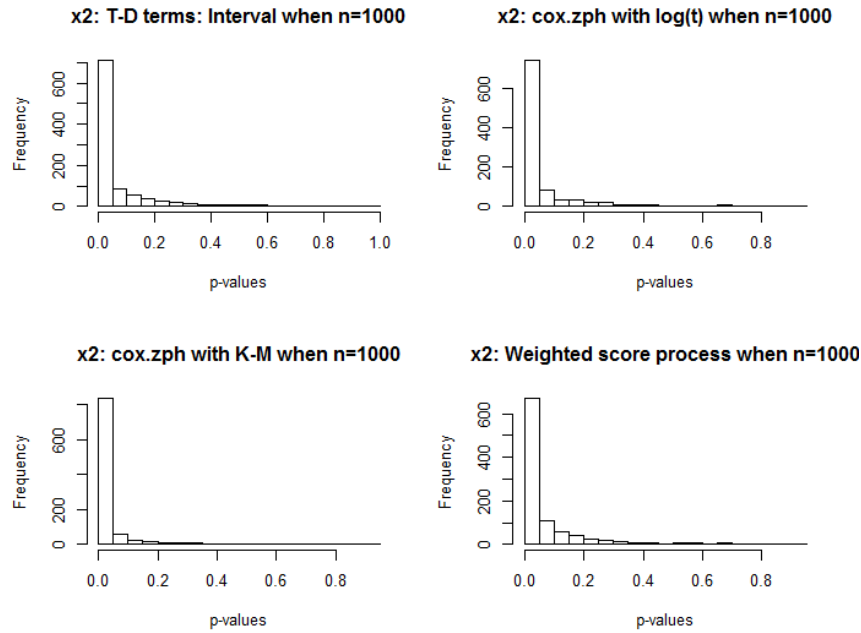
**Figure 4.5** Histograms of  $p$ -values for  $x_2$  when  $c = -0.3$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



**Figure 4.6** Histograms of  $p$ -values for  $x_2$  when  $c = -0.3$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



**Figure 4.7** Histograms of  $p$ -values for  $x_2$  when  $c = -0.6$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



**Figure 4.8** Histograms of  $p$ -values for  $x_2$  when  $c = -0.6$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.

respectively. For the same tests with  $c = -0.6$ , the histograms of  $p$ -values for  $x_2$  are given in Figure 4.7 and Figure 4.8 for  $n = 250$  and  $n = 1000$  individuals, respectively.

In the case of  $c = 0.3$  and  $c = 0.6$ , it follows from Table 4.3 that the “Weighted score process” does not achieve the correct level for  $x_1$  in three of the four cases with respectively, 3.0%, 3.9% and 7.8%. What is worth mentioning here is when  $c = 0.6$  with  $n = 1000$  the achieved significance level for  $x_1$  in those of the tests are quite large, and only the test “cox.zph with  $\log t$ ” and “cox.zph with  $\hat{S}(t)$ ”, with respectively 5.9% and 5.8%, are close to the nominal 5% level. Further, it follows from the table that the test “cox.zph with  $\hat{S}(t)$ ” obtained the greatest value of the achieved power for  $x_2$  when  $n = 1000$  for both  $c = 0.3$  and  $c = 0.6$  with 46.1% and 96.3%, respectively. The histograms of  $p$ -values for  $x_2$  from the test “T-D terms: Interval”, “cox.zph with  $\log t$ ”, “cox.zph with  $\hat{S}(t)$ ” and “Weighted score process” illustrating the results with  $c = 0.3$  are given in Figure A.12 and Figure A.13 for  $n = 250$  and  $n = 1000$  individuals, respectively, in Appendix A. In case of  $c = 0.6$ , the histograms of  $p$ -values for the same tests are shown in Figure A.14 and Figure A.15 for  $n = 250$  and  $n = 1000$  individuals, respectively.

To sum up, based on the achieved power from the tests for  $x_2$ , in the case of  $c = -0.3$  and  $c = -0.6$ , the test “cox.zph with  $\hat{S}(t)$ ” obtained the most optimal results and therefore the most efficient test of non-proportionality. In the situation with  $c = 0.3$  and  $c = 0.6$ , a comparison among the tests shows that the “cox.zph with  $\hat{S}(t)$ ” is still the most efficient tests of non-proportional assumption.

#### 4.3.3 Another case of model with time-varying coefficients

We will still consider the incorrectly specified model (4.7), but instead of (4.8), consider the time-varying coefficient of the following form

$$\beta_2(t) = \begin{cases} \beta_2 + c, & \text{for } t < t_0 \\ \beta_2 - c, & \text{for } t \geq t_0, \end{cases} \quad (4.24)$$

where  $c \geq 0$  and  $t_0 > 0$  are constants. Note that  $c = 0$  corresponds to the proportional hazard model.

##### When the covariate $x_{i2} = 0$

Since  $x_{i2}$  is Bernoulli distributed, the survival time  $T_i$  for an individual  $i$  when  $x_{i2} = 0$  is exactly the same one as given in (4.9).

##### When the covariate $x_{i2} = 1$

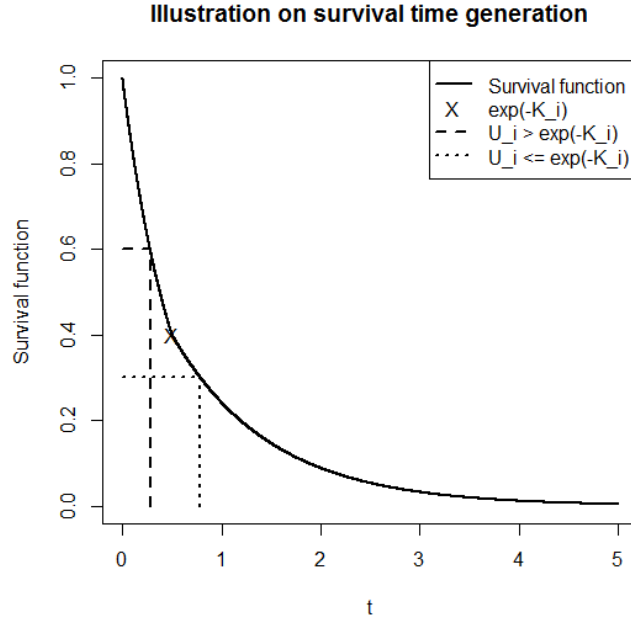
When  $x_{i2} = 1$ , the hazard rate for an individual  $i$  is given by

$$\alpha(t|x_{i1}) = \begin{cases} (a/b^a)t^{a-1} \exp\{\beta_1 x_{i1} + \beta_2 + c\}, & \text{for } t < t_0 \\ (a/b^a)t^{a-1} \exp\{\beta_1 x_{i1} + \beta_2 - c\}, & \text{for } t \geq t_0. \end{cases} \quad (4.25)$$

The corresponding cumulative hazard rate is formulated in this way:

- If  $t < t_0$ , then

$$A(t|x_{i1}) = (t/b)^a \exp\{\beta_1 x_{i1} + \beta_2 + c\}. \quad (4.26)$$



**Figure 4.9** For a given value of  $x_{i1}$ , this is an illustration on how the survival time  $T_i$  for an individual  $i$  is generated when  $x_{i2} = 1$ . For instance, if  $U_i = 0.6$ , then the corresponding survival time  $T_i = 0.278$ . On the other hand, if  $U_i = 0.3$ , then  $T_i = 0.785$ . Note that  $K_i$  is given by (4.28).

- If  $t \geq t_0$ , then

$$\begin{aligned}
 A(t|x_{i1}) &= (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2 + c\} \int_0^{t_0} u^{a-1} du \\
 &\quad + (a/b^a) \exp\{\beta_1 x_{i1} + \beta_2 - c\} \int_{t_0}^t u^{a-1} du \\
 &= (t_0/b)^a \exp\{\beta_1 x_{i1} + \beta_2 + c\} + \left[ (t/b)^a - (t_0/b)^a \right] \exp\{\beta_1 x_{i1} + \beta_2 - c\}.
 \end{aligned} \tag{4.27}$$

When  $t = t_0$ , the cumulative hazard rate for an individual  $i$  obtains the value

$$K_i = A(t_0|x_{i1}) = (t_0/b)^a \exp\{\beta_1 x_{i1} + \beta_2 + c\}. \tag{4.28}$$

As illustrated in Figure 4.9, to generate the survival time  $T_i$ , we first generate  $U_i$  from the  $U(0, 1)$ -distribution, and then obtain  $T_i$  as follows:

- If  $U_i > \exp\{-K_i\}$ , then by using (4.26), we obtain

$$T_i = \left( -b^a \log(U_i) \exp\{-\beta_1 x_{i1} - \beta_2 - c\} \right)^{1/a}. \tag{4.29}$$

- If  $U_i \leq \exp\{-K_i\}$ , then by using (4.27), we obtain

$$T_i = \left( [1 - \exp(2c)]t_0^a - b^a \log(U_i) \exp\{-\beta_1 x_{i1} - \beta_2 + c\} \right)^{1/a}. \tag{4.30}$$

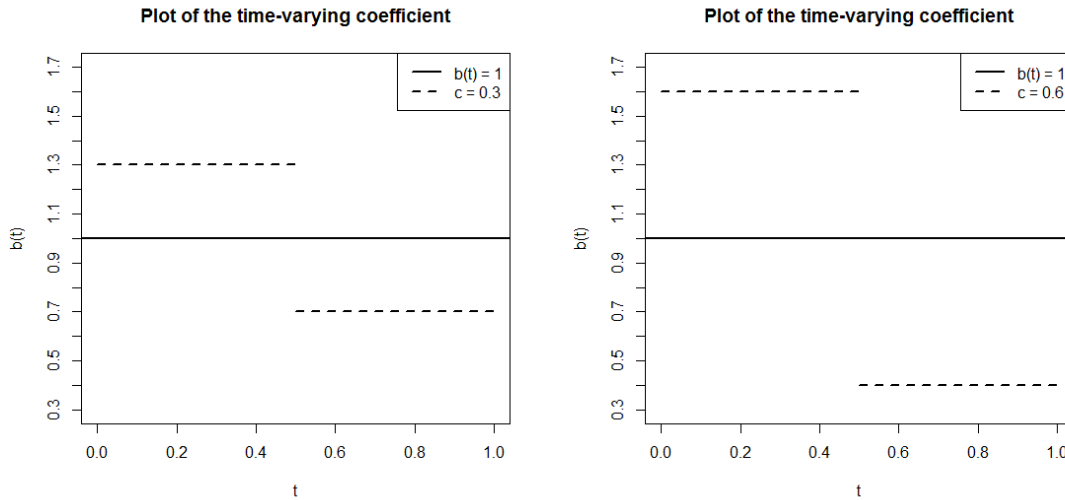
To sum up, the survival time  $T_i$  for the  $i$ th individual is generated from (4.9) when  $x_{i2} = 0$ . When  $x_{i2} = 1$ , then  $T_i$  is generated from (4.29) if  $U_i > \exp\{-K_i\}$  holds, else it is generated from (4.30).

#### 4.3.4 Check the proportionality of the incorrectly specified model

We will use the same simulation procedure as mentioned in the previous illustration, but the survival time  $T_i$  will be generated from (4.9), (4.29) and (4.30) instead. Let  $\beta_1 = 0.5$ ,  $\beta_2 = 1$ ,  $\tau = 1$ , the shape parameter  $a = 1$  and the scale parameter  $b = 2$ . The covariate  $x_{i1}$  and  $x_{i2}$  are generated from  $N(0,1)$  and  $\text{Bernoulli}(0.5)$ , respectively. From now on, the constant  $c$  is the level relative to  $\beta_2$ , while  $t_0$  is a crucial point that affect the function curve of  $\beta_2(t)$  in (4.24) to shift the level from positive to negative. We will take a look at the case when  $c = 0.3$  and  $c = 0.6$ , with  $t_0 = 0.5$  in both cases. A plot of  $\beta_2(t)$  vs. time  $t \in [0, 1]$  is shown in Figure 4.10, where the figure on the left-hand side corresponding to the case with  $c = 0.3$ .

Note that the censoring time  $C_i$ , censored survival time  $\tilde{T}_i$  and censoring indicator  $D_i$  are generated in the same way as mentioned in Subsection 4.2.1. Further, the event rate will still be kept at around 50%, and a reasonable choice of the censoring rate for  $c = 0.3$  and  $c = 0.6$  are  $\lambda = 0.920$  and  $\lambda = 1.150$ , respectively.

Since  $x_1$  has a proportional effect, we want to obtain an achieved significance level close to the nominal 5% level. From Table 4.4, we see that when  $c = 0.3$  and  $n = 250$ , the “T-D terms: Interval”, “cox.zph with  $\hat{S}(t)$ ” and “Weighted score process” don’t achieves the correct level with respectively, 3.3%, 3.6% and 2.5%. When  $c = 0.6$  and  $n = 250$ , the “Weighted score process” is the only one that does not obtain this criteria with 3.1%, while with  $n = 1000$  the test “T-D terms: log  $t$ ” and “Weighted score process” don’t obtain this criteria with respectively, 8.1% and 8.9%.



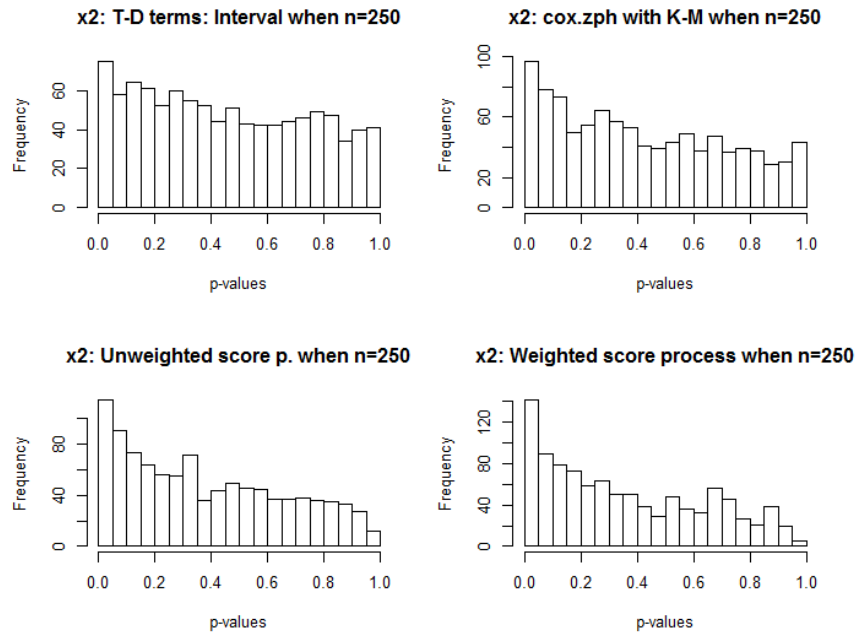
**Figure 4.10** A plot of the time-varying coefficient  $\beta_2(t)$  vs. time  $t \in [0, 1]$ . Left panel: When  $c = 0.3$ . Right panel: When  $c = 0.6$ . At time  $t = 0.5$  both curves shift the level from positive to negative. The horizontal line corresponding to the assumption of proportionality.

**Table 4.4** Achieved significance level and achieved power for  $x_1$  and  $x_2$ , respectively, in the case of  $c = 0.3$  and  $c = 0.6$  over 1000 simulated data sets of survival times for both  $n = 250$  and  $n = 1000$  individuals. The Cox regression model (4.7) is incorrectly specified with  $\boldsymbol{\beta} = (0.5, 1)^T$ .

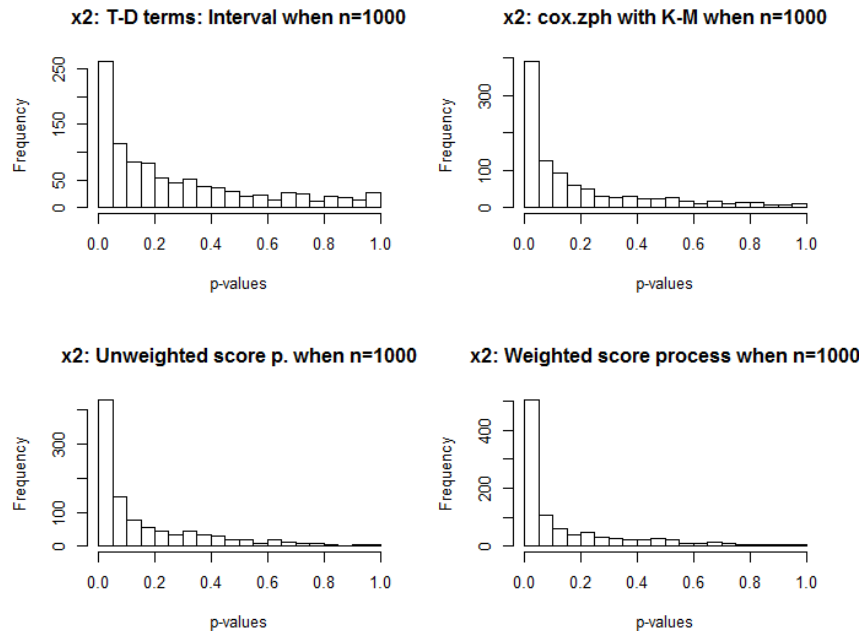
Test	For $c = 0.3$				For $c = 0.6$			
	$n = 250$		$n = 1000$		$n = 250$		$n = 1000$	
	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$
T-D terms: $\log t$	4.1%	7.6%	5.6%	24.1%	5.2%	16.8%	8.1%	48.7%
T-D terms: Interval	3.3%	7.5%	5.1%	26.4%	5.9%	11.7%	6.5%	34.6%
cox.zph with $\log t$	4.6%	6.7%	5.3%	22.3%	5.6%	12.9%	6.9%	41.3%
cox.zph with $\hat{S}(t)$	3.6%	9.7%	5.5%	39.2%	5.1%	21.8%	5.5%	65.0%
Unweighted score process	4.6%	11.5%	5.2%	42.9%	5.6%	23.8%	6.0%	76.8%
Weighted score process	2.5%	14.2%	4.6%	50.5%	3.1%	35.0%	8.9%	90.6%
cox.zph Global test: $\log t$	5.9%		17.3%		10.0%		34.9%	
cox.zph Global test: $\hat{S}(t)$	7.3%		30.2%		17.9%		58.6%	

In terms of  $x_2$ , a comparison among the tests shows that when  $c = 0.3$  and  $c = 0.6$ , the weakest test of non-proportionality are, respectively, the “cox.zph with  $\log t$ ” and “T-D terms: Interval”. On the other hand, the “Weighted score process” obtained the greatest value of the achieved power for  $x_2$  in all cases. The conclusion here is that the “Weighted score process” has both good and bad properties when testing the proportionality assumption of the incorrectly specified model. The good one is when we are testing the covariate  $x_2$  which has a non-proportional effect, then the test will achieve the most optimal results compared with the other five tests. The bad one is when we are testing the covariate  $x_1$  which has a proportional effect, then the test will not achieve the nominal 5% level in most of the situations. Thus, what we should prefer in this context may be the “Unweighted score process”, which gave us the most satisfactory results of the achieved significance level and the achieved power for  $x_1$  and  $x_2$ , respectively.

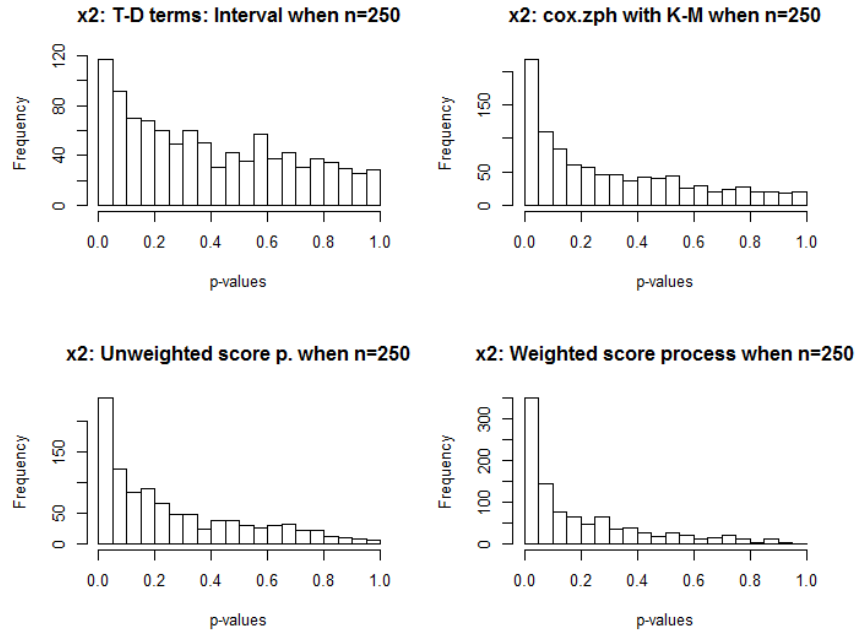
Histograms of  $p$ -values for  $x_2$  based on the test “T-D terms: Interval”, “cox.zph with  $\hat{S}(t)$ ”, “Unweighted score process” and “Weighted score process” when  $c = 0.3$  are given in Figure 4.11 and Figure 4.12 for  $n = 250$  and  $n = 1000$  individuals, respectively. What is worth mentioning here is when the number of individuals increases from  $n = 250$  to  $n = 1000$ , the number of non-proportionality cases for  $x_2$  also increases, which is clearly shown in the plot. This fact is more obvious when  $c = 0.6$ , and for the same tests are shown in Figure 4.13 and Figure 4.14 for  $n = 250$  and  $n = 1000$  individuals, respectively. Other plots for the same tests, but for  $x_1$  are moved to Appendix A.



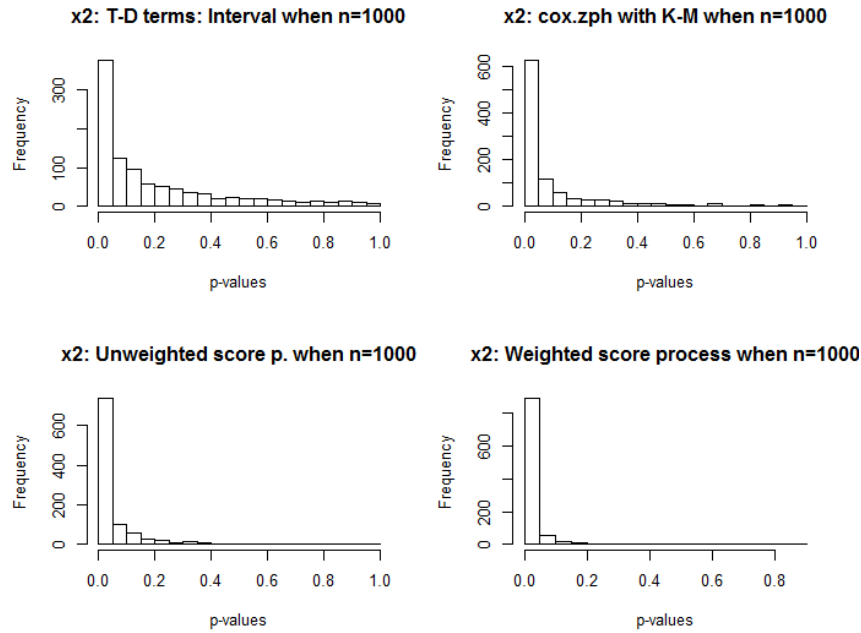
**Figure 4.11** Histograms of  $p$ -values for  $x_2$  when  $c = 0.3$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



**Figure 4.12** Histograms of  $p$ -values for  $x_2$  when  $c = 0.3$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



**Figure 4.13** Histograms of  $p$ -values for  $x_2$  when  $c = 0.6$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



**Figure 4.14** Histograms of  $p$ -values for  $x_2$  when  $c = 0.6$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



## Chapter 5

# Concluding remarks

Cox's regression model is one of the most used methods in medical statistics, as well as applications in other fields. The proportional hazards assumption is one of the crucial assumptions in this method that have to be satisfied. It is therefore important to know how this can be checked before using the model in practice, which is the purpose of the current thesis.

In Chapter 3, we have presented two models, (3.4) and (3.5), that extend the Cox regression model with one or more time-dependent terms, and checked the proportional hazards assumption by using the score test in (2.33). We have also presented the tests based on the scaled Schoenfeld residual and the tests based on the score process for checking this assumption. As an illustration, a real data set from the German Breast Cancer Study Data have been used for checking the performance of the tests. Further, it follows from the calculated  $p$ -values that the covariate Tumor Grade 3 has a non-proportional effect. The test of the model (3.4) and the "Unweighted score process" shows also that the log-transformed Number of Progesterone Receptor has a non-proportional effect, which is not the same case for the other tests.

In Chapter 4, simulation have been used to generate our own data set of survival times under a variety of situations for non-proportional hazards. We have used two covariate,  $x_1$  and  $x_2$ , to fit the Cox regression model, where  $x_1$  is standard normally distributed, while  $x_2$  is Bernoulli distributed with success probability  $p = 0.5$ . Further, we have performed 1000 simulations for both 250 and 1000 individuals in the study, and performed the tests of the proportionality assumption under each simulation. Thus, when the model is correctly specified, the achieved significance level for both covariates should be close to the nominal 5% level. According to the results, the "Weighted score process" is the only one that does not achieve the correct level.

For the incorrectly specified model in Chapter 4, we have reformulated the coefficient for  $x_2$  to be time-varying. That is, since  $x_2$  has a non-proportional effect, the assumption of proportionality should be rejected in most of the cases in the simulations. This corresponds to a large value of the achieved power for those tests. It follows from the results that when the time-varying coefficient has a gradually deviates from the proportionality, as given in (4.8) and shown in Figure 4.4, then the "cox.zph with  $\hat{S}(t)$ " is the most efficient test of the proportional and non-proportional effect for  $x_1$  and  $x_2$ , respectively. On the other hand, when the time-varying coefficient has a sudden deviates from the proportionality, as given in (4.24) and shown in Figure 4.10, then the "Weighted score process" is the most

efficient test of non-proportional effect. Note, however, that this test at the same time is also the weakest one for checking the proportional effect of  $x_1$ . In this case, the test that obtained the most satisfactory results for both the achieved significance level and the achieved power for  $x_1$  and  $x_2$ , respectively, is the “Unweighted score process”.

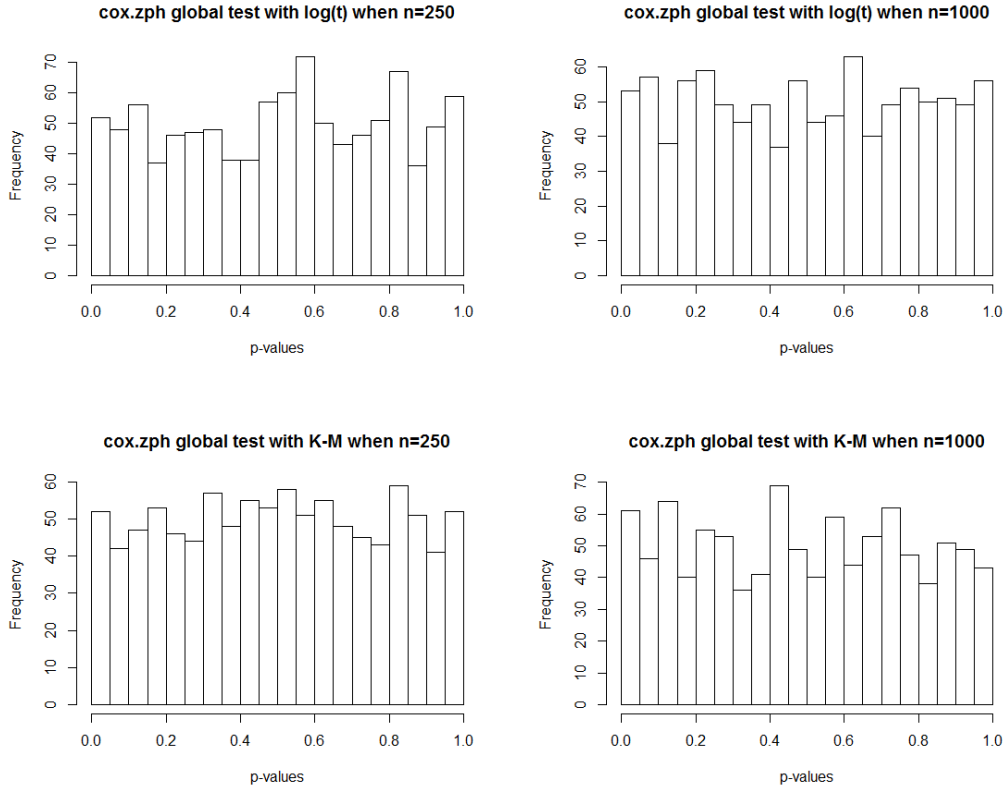
To conclude, in case of a sudden change of level in the hazard function, it is recommended to use the “Unweighted score process” for checking the proportionality of the Cox model. Otherwise the method which uses the implementation `cox.zph` with default time-transform in **R** is preferred.

Due to limited time we are not able to investigate the case when  $x_1$  and  $x_2$  are correlated. The case where both covariates have a non-proportional effect is not included neither. Further research could also contain a study focusing on modifications of the existing test, e.g. the unweighted score process of Lin et al. (1993), where the existing test uses the supremum of the absolute value of the score process as test statistic, but were other measures of deviation from zero could be more relevant. In addition, we have eased out our study to include two covariates only, but a third or more other covariates could also be considered. Note however that our main focus is not to find the optimal model, but to give a guidance on which test performs best under different circumstances. Moreover, it could also be interesting to check whether using more simulations could give more accuracy or better performance of the tests. Last but not least, one could also try different initial values of the parameters such as  $\tau$  and  $t_0$  and see whether the choice of these parameters has significant impact on the final results.

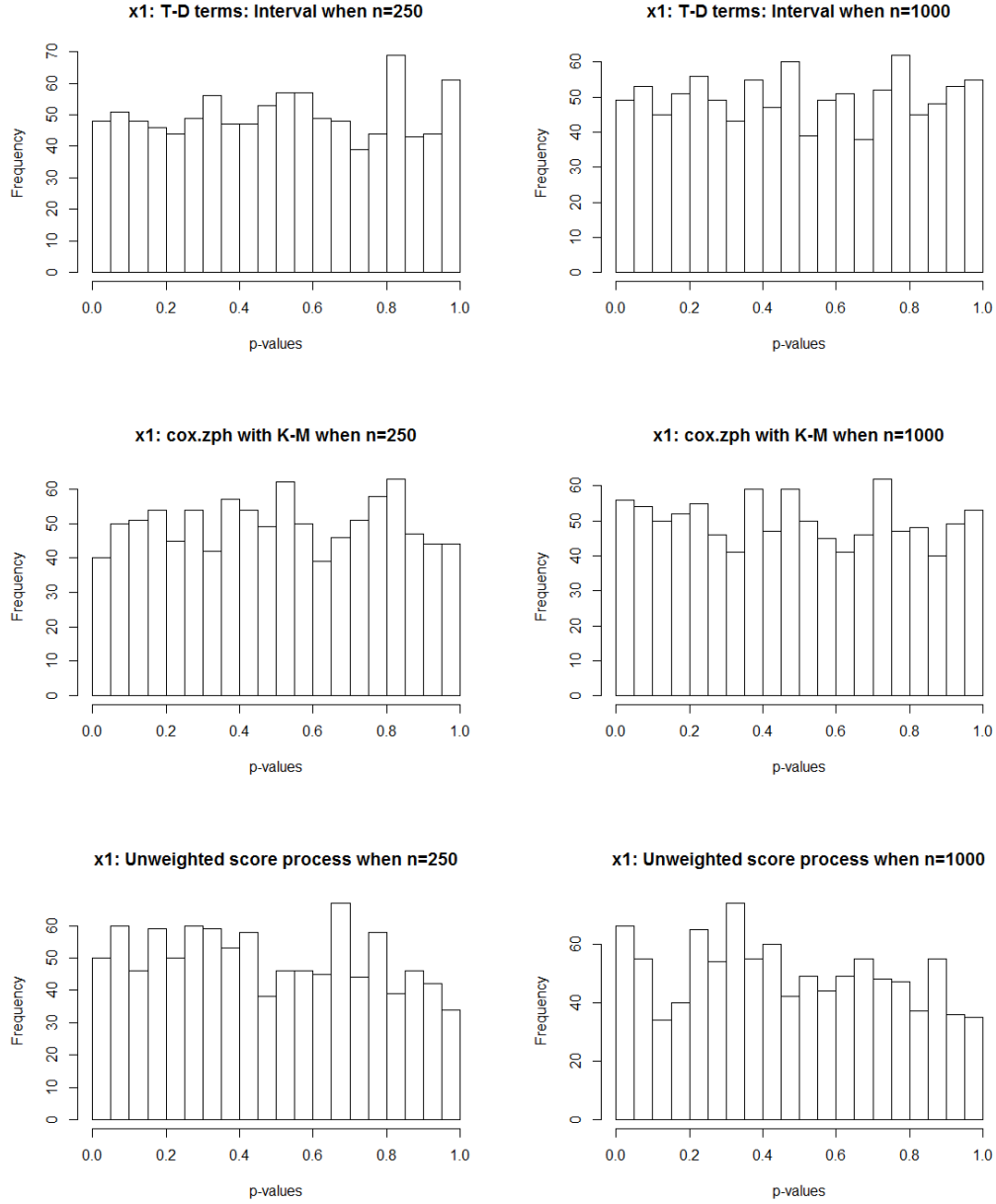
# Appendix A

## Figures

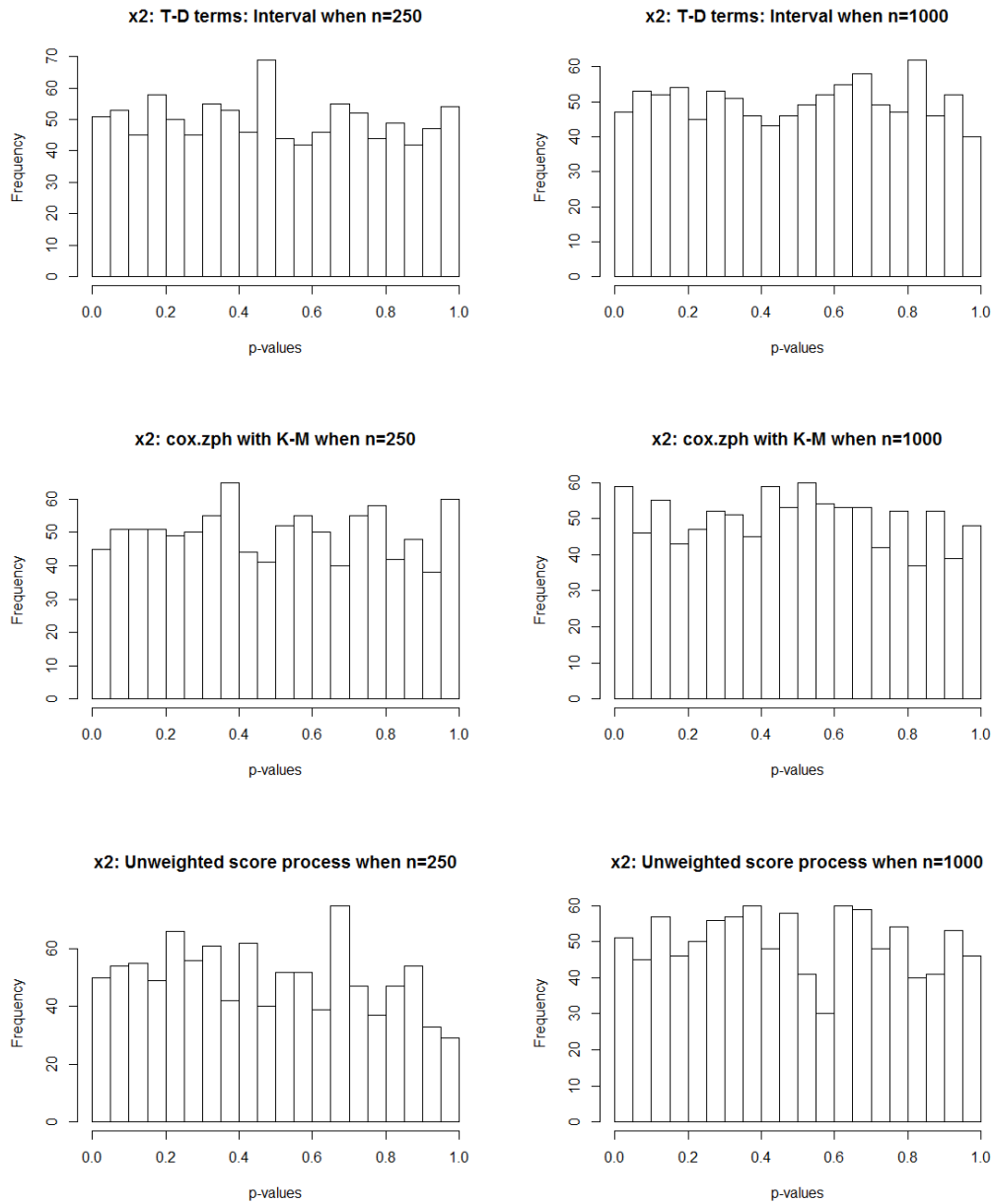
The following section contains a number of histogram plots of the  $p$ -values from the simulation part in Subsection 4.2.2, 4.3.2 and 4.3.4. These plots have similar results, and therefore not to be included in the illustration part. The subtext under every plots tells us which section and cases they belongs to.



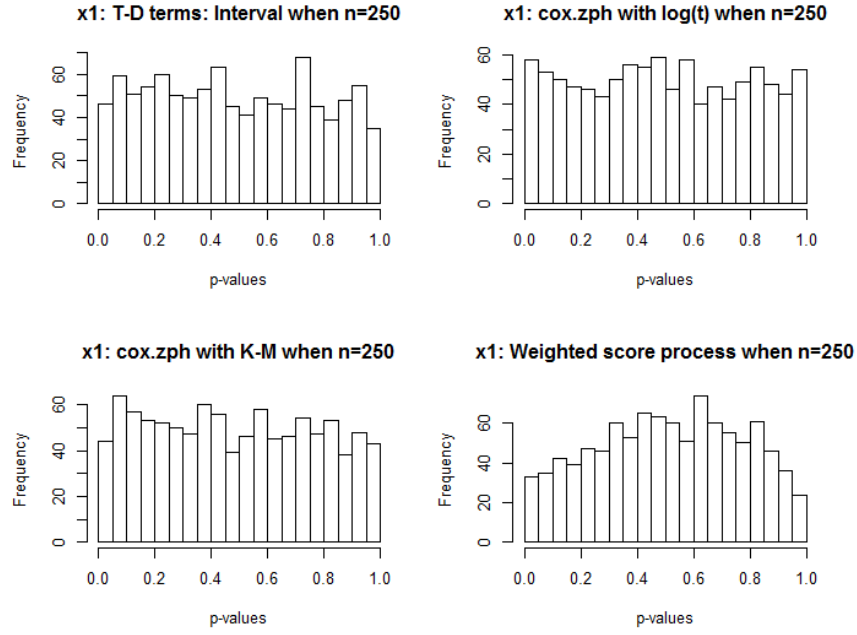
**Figure A.1** From Subsection 4.2.2: Histograms of  $p$ -values over 1000 simulated data sets of survival times for  $n = 250$  individuals (left panels) and  $n = 1000$  individuals (right panels) based on the global test of `cox.zph` function in **R** with  $\log t$  and  $\hat{S}(t)$  time-transform function when the model is correctly specified with  $\boldsymbol{\beta} = (0.5, 1)^T$ .



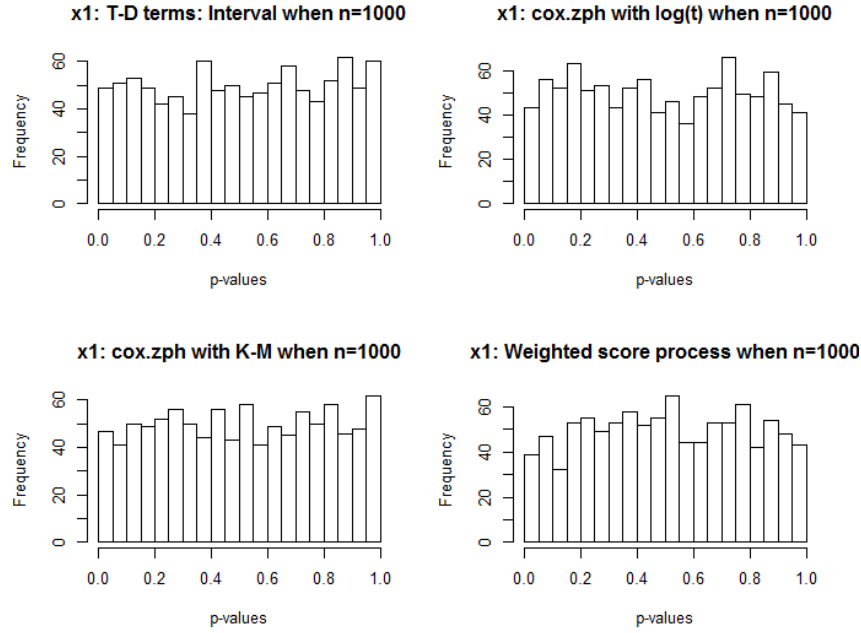
**Figure A.2** From Subsection 4.2.2: Histograms of p-values for  $x_1$  over 1000 simulated data sets of survival times for  $n = 250$  individuals (left panels) and  $n = 1000$  individuals (right panels) based on three different tests of proportionality when the model is correctly specified with  $\beta = (0.5, 1)^T$ .



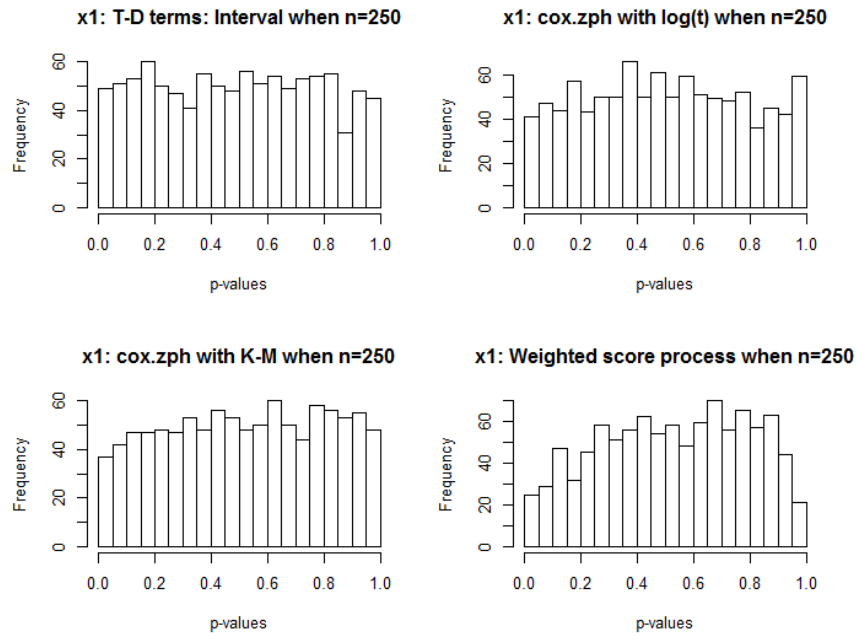
**Figure A.3** From Subsection 4.2.2: Histograms of p-values for  $x_2$  over 1000 simulated data sets of survival times for  $n = 250$  individuals (left panels) and  $n = 1000$  individuals (right panels) based on three different tests of proportionality when the model is correctly specified with  $\beta = (0.5, 1)^T$ .



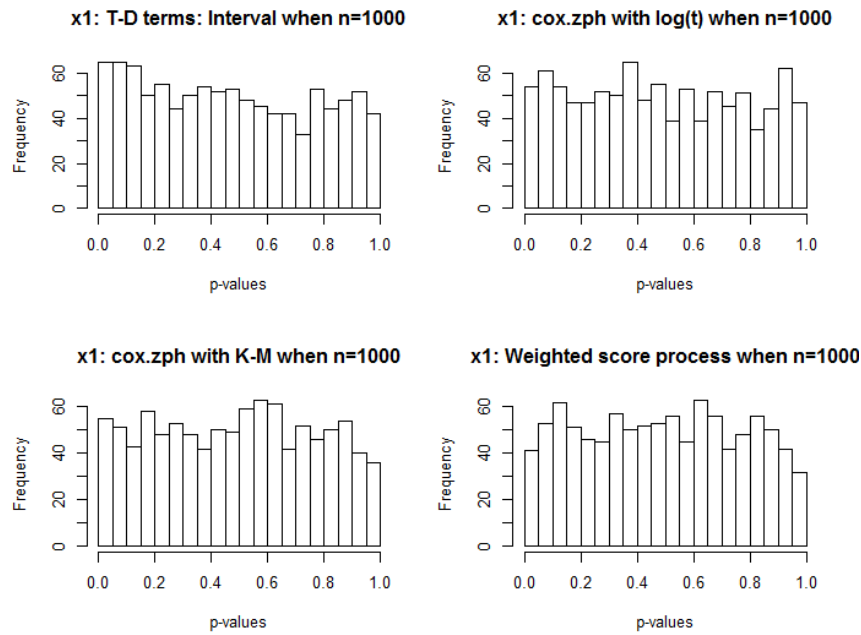
**Figure A.4** From Subsection 4.3.2: Histograms of p-values for  $x_1$  when  $c = -0.3$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



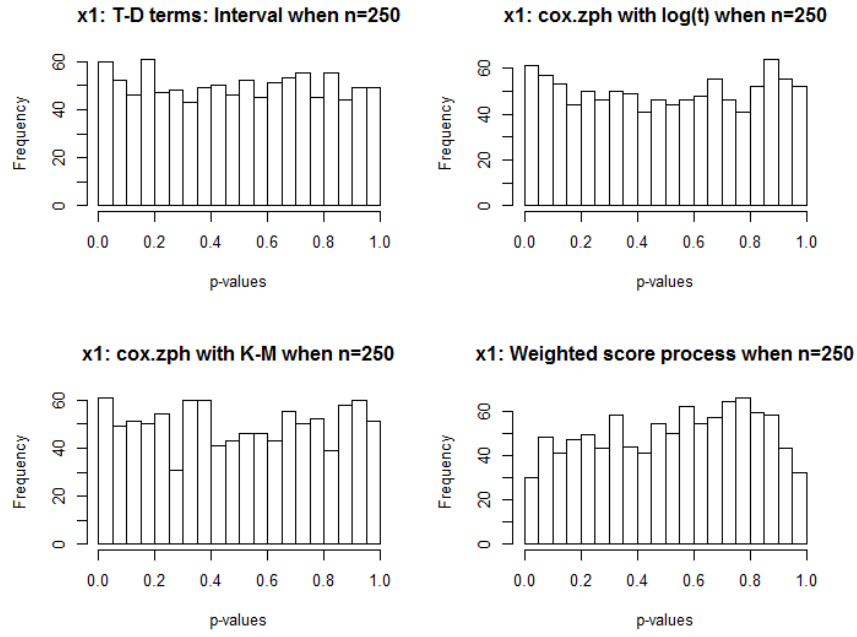
**Figure A.5** From Subsection 4.3.2: Histograms of p-values for  $x_1$  when  $c = -0.3$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



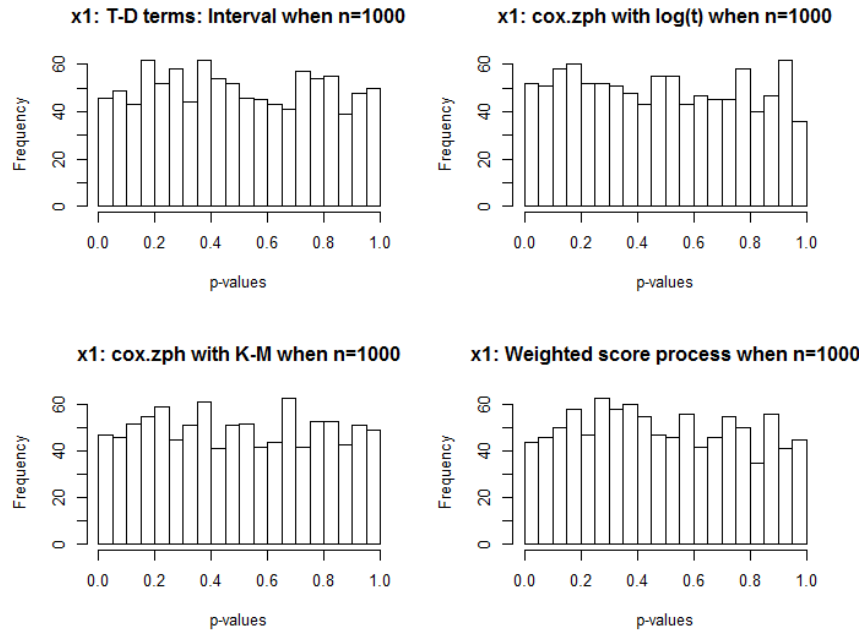
**Figure A.6** From Subsection 4.3.2: Histograms of p-values for  $x_1$  when  $c = -0.6$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



**Figure A.7** From Subsection 4.3.2: Histograms of p-values for  $x_1$  when  $c = -0.6$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.

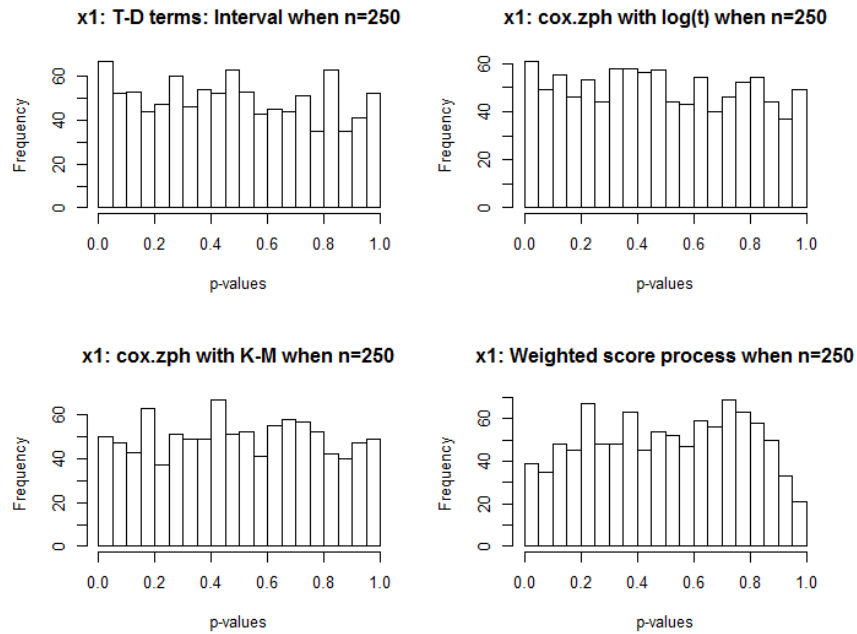


**Figure A.8** From Subsection 4.3.2: Histograms of  $p$ -values for  $x_1$  when  $c = 0.3$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.

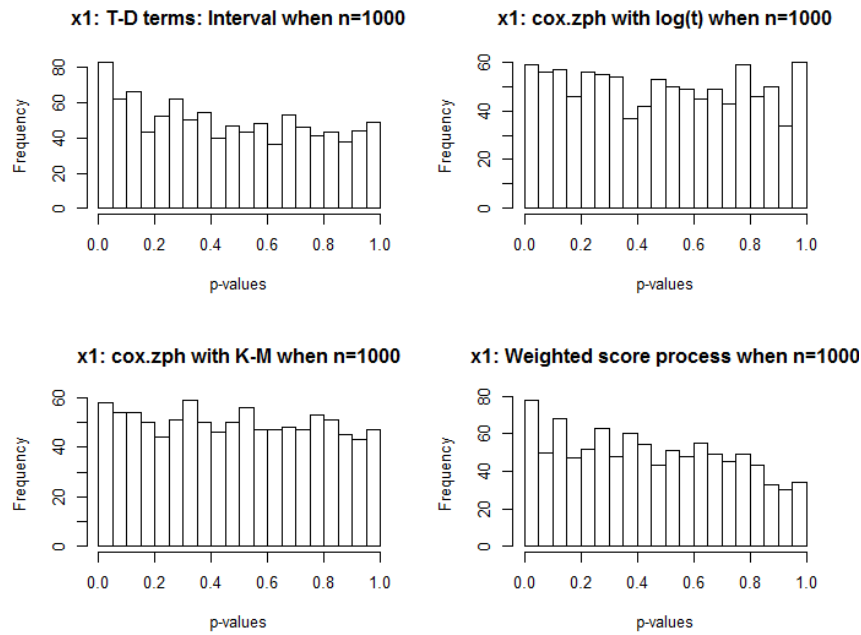


**Figure A.9** From Subsection 4.3.2: Histograms of  $p$ -values for  $x_1$  when  $c = 0.3$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.

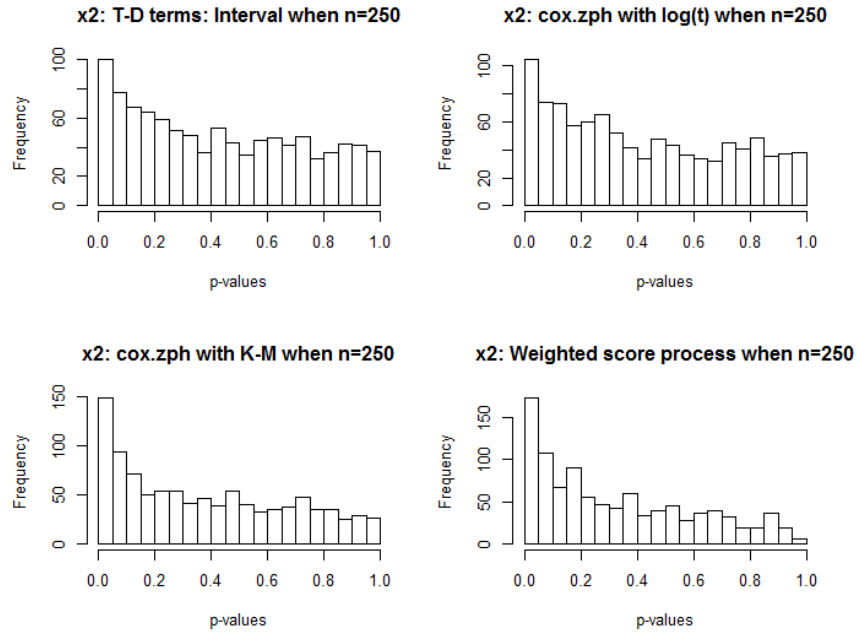




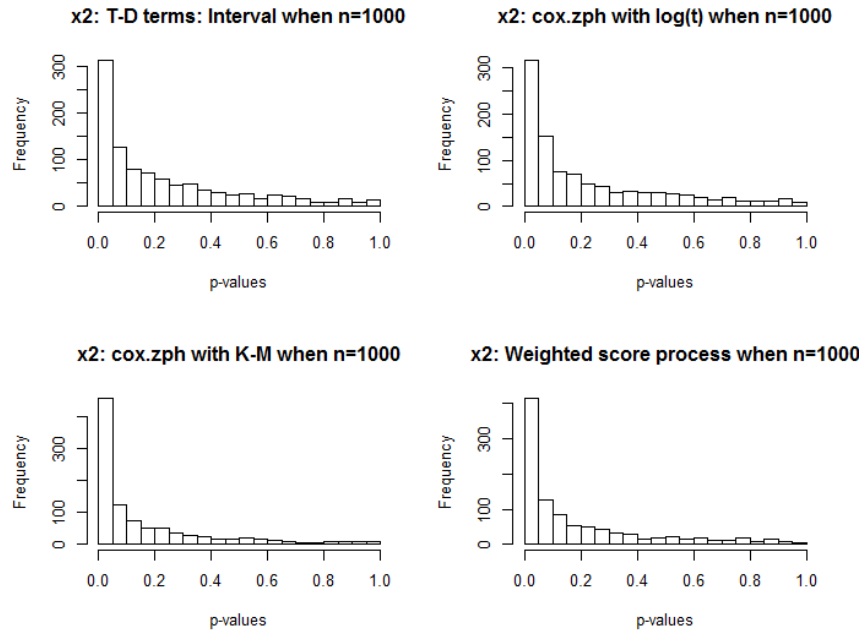
**Figure A.10** From Subsection 4.3.2: Histograms of p-values for  $x_1$  when  $c = 0.6$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



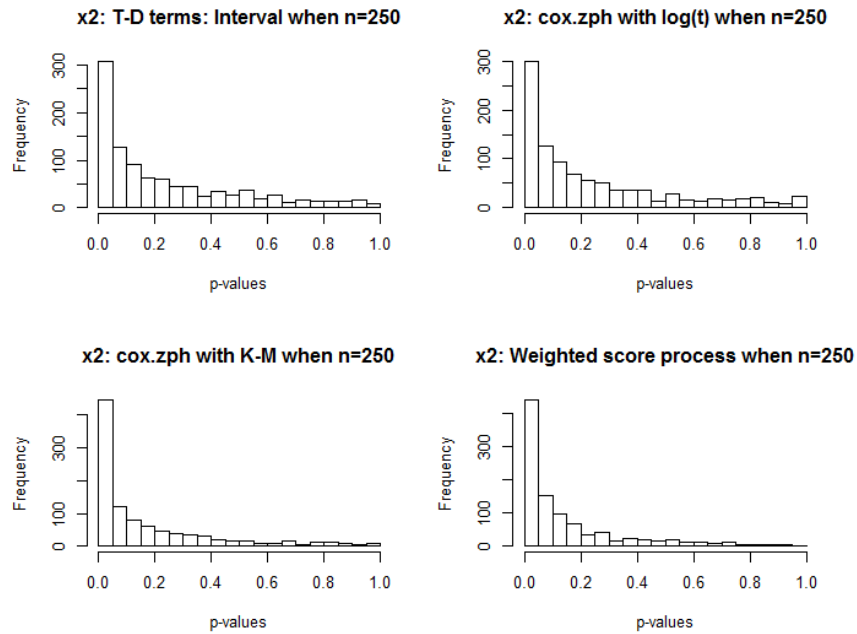
**Figure A.11** From Subsection 4.3.2: Histograms of p-values for  $x_1$  when  $c = 0.6$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



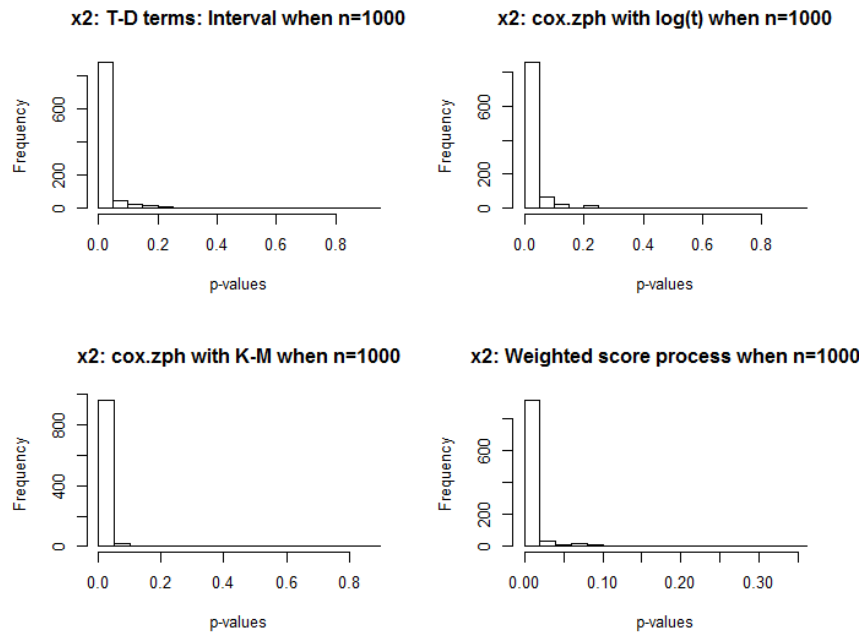
**Figure A.12** From Subsection 4.3.2: Histograms of p-values for  $x_2$  when  $c = 0.3$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



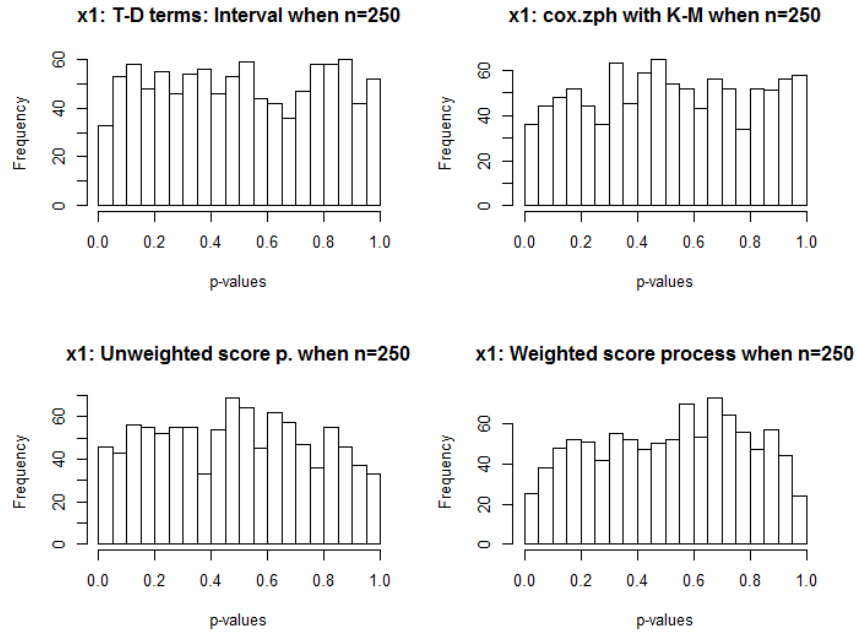
**Figure A.13** From Subsection 4.3.2: Histograms of p-values for  $x_2$  when  $c = 0.3$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



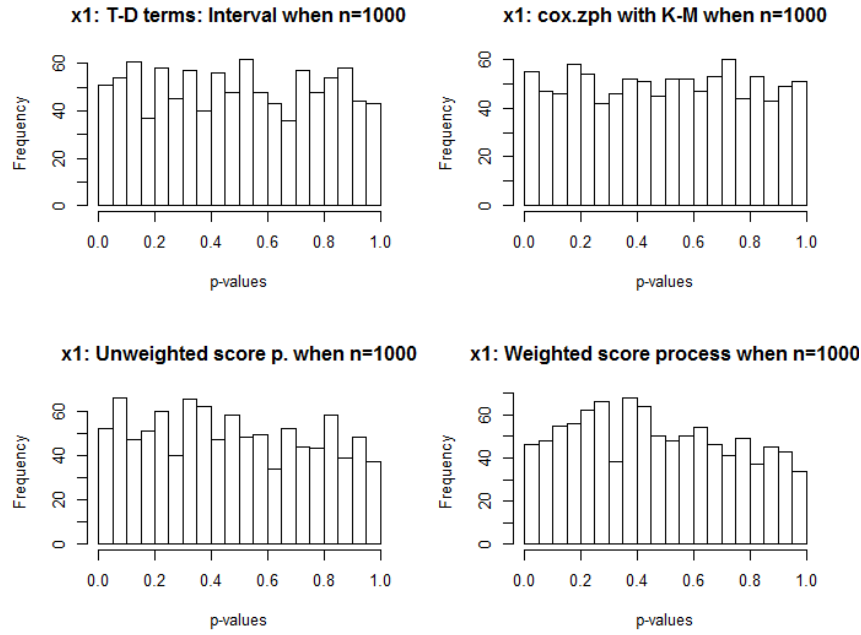
**Figure A.14** From Subsection 4.3.2: Histograms of p-values for  $x_2$  when  $c = 0.6$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



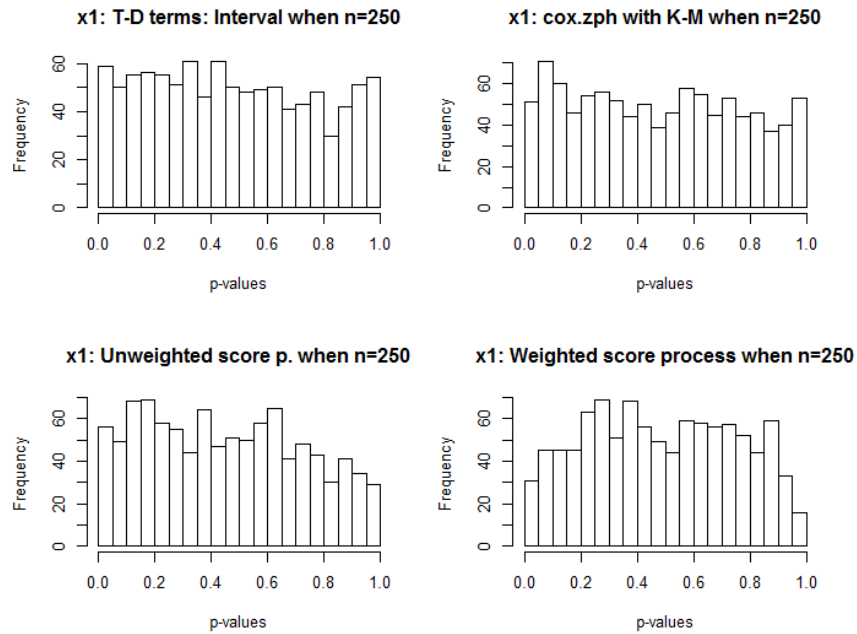
**Figure A.15** From Subsection 4.3.2: Histograms of p-values for  $x_2$  when  $c = 0.6$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



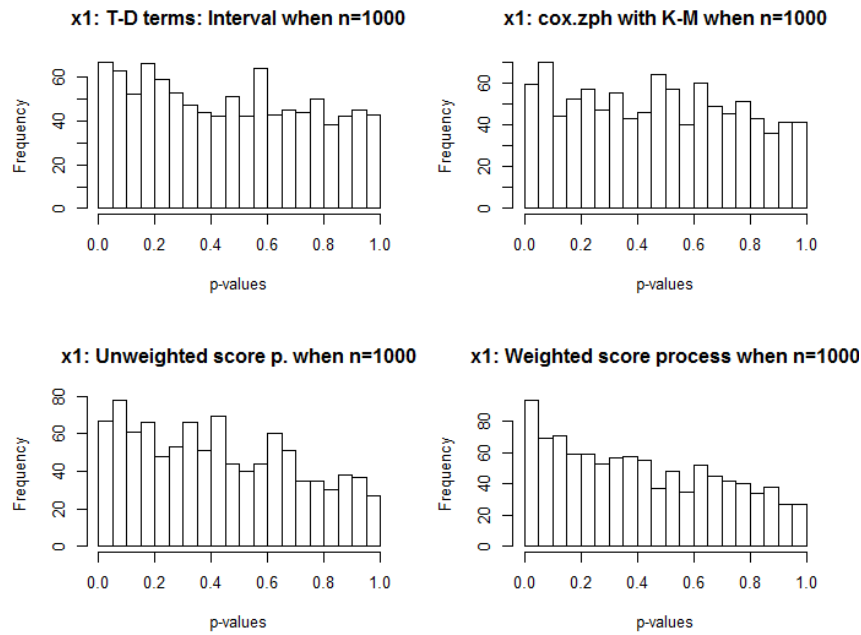
**Figure A.16** From Subsection 4.3.4: Histograms of  $p$ -values for  $x_1$  when  $c = 0.3$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



**Figure A.17** From Subsection 4.3.4: Histograms of  $p$ -values for  $x_1$  when  $c = 0.3$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



**Figure A.18** From Subsection 4.3.4: Histograms of p-values for  $x_1$  when  $c = 0.6$  and  $n = 250$  individuals over 1000 simulated data sets of survival times.



**Figure A.19** From Subsection 4.3.4: Histograms of p-values for  $x_1$  when  $c = 0.6$  and  $n = 1000$  individuals over 1000 simulated data sets of survival times.



# Bibliography

Aalen, O. O., Borgan, Ø. and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer-Verlag, New York.

Bender, R., Augustin, T. and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* **24**, 1713-1723.

Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **34**, 187-220.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515-526.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data* (2nd ed.). Springer-Verlag, New York.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-572.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, <http://www.R-project.org>.

Sauerbrei, W. and Royston, P. (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society, Series A* **162**, 71-94.

Therneau, T. M. and Grambsch P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

Van Noorden, R., Maher, B. and Nuzzo, R. (2014). The top 100 papers. *Nature* **514**, 550-553.